

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



# California Alternate Assessment for Science 2018–‍2019 Field Test Technical Report

**Prepared for the California Department of Education by Educational Testing Service**

**Submitted June 1, 2020**



**Contract No. CN150012**

Table of Contents

[Chapter 1: Introduction 1](#_Toc180396433)

[1.1. Background 1](#_Toc180396434)

[1.2. Purpose of the Field Test 2](#_Toc180396435)

[1.3. Field Test Content and Design 3](#_Toc180396436)

[1.3.1. Assessment Model 3](#_Toc180396437)

[1.3.2. California Next Generation Science Standards Core Content Connectors (Science Connectors) 4](#_Toc180396438)

[1.3.3. Test Components for the Field Test 4](#_Toc180396439)

[1.4. Intended Population 5](#_Toc180396440)

[1.5. Testing Window and Times 5](#_Toc180396441)

[1.6. Significant Developments for the CAA for Science 2018–2019 Field Test 6](#_Toc180396442)

[1.6.1. Online Delivery 6](#_Toc180396443)

[1.7. Limitations of the Field Test 6](#_Toc180396444)

[1.8. Groups and Organizations Involved with the Assessment 6](#_Toc180396445)

[1.8.1. State Board of Education (SBE) 6](#_Toc180396446)

[1.8.2. California Department of Education (CDE) 6](#_Toc180396447)

[1.8.3. California Educators 7](#_Toc180396448)

[1.8.4. Contractors 7](#_Toc180396449)

[1.9. Systems Overview and Functionality 8](#_Toc180396450)

[1.9.1. Test Operations Management System (TOMS) 8](#_Toc180396451)

[1.9.2. Test Delivery System (TDS) 8](#_Toc180396452)

[1.9.3. Training Test 8](#_Toc180396453)

[1.9.4. Online Reporting System (ORS) and California Educator Reporting System (CERS) 9](#_Toc180396454)

[1.10. Overview of the Technical Report 9](#_Toc180396455)

[References 11](#_Toc180396456)

[Chapter 2: Overview of the CAA for Science Field Test Processes 12](#_Toc180396457)

[2.1. Embedded Performance Task (PT) and Item Development and Review 12](#_Toc180396458)

[2.1.1. Selection of Science Connectors for Embedded PT Development 12](#_Toc180396459)

[2.1.2. Embedded PT Development for Grades Five and Eight and High School 12](#_Toc180396460)

[2.1.3. Universal Design Principles 13](#_Toc180396461)

[2.2. Test Assembly 14](#_Toc180396462)

[2.2.1. Test Design 14](#_Toc180396463)

[2.2.2. Test Blueprints 15](#_Toc180396464)

[2.2.3. Test Length 15](#_Toc180396465)

[2.3. Test Administration 15](#_Toc180396466)

[2.3.1. Test Security and Confidentiality 15](#_Toc180396467)

[2.3.2. Procedures to Maintain Standardization 16](#_Toc180396468)

[2.4. Test-Taking Rates 16](#_Toc180396469)

[2.5. Fairness and Accessibility 17](#_Toc180396470)

[2.5.1. Universal Tools, Designated Supports, and Accommodations 17](#_Toc180396471)

[2.5.2. Individualizations 19](#_Toc180396472)

[2.6. Scores 19](#_Toc180396473)

[2.6.1. Score Reporting 19](#_Toc180396474)

[2.6.2. Aggregation Procedures 19](#_Toc180396475)

[2.7. Overview of Psychometric Analyses 19](#_Toc180396476)

[2.7.1. Description of the Classical Item Analyses 19](#_Toc180396477)

[2.7.2. Description of DIF Analyses 20](#_Toc180396478)

[2.7.3. Description of Item Response Theory (IRT) Analyses 20](#_Toc180396479)

[References 21](#_Toc180396480)

[Appendix 2.A: Test-Taking Rates 22](#_Toc180396481)

[Chapter 3: Embedded PT Development and Review 26](#_Toc180396482)

[3.1. Embedded PT and Item Development 26](#_Toc180396483)

[3.1.1. Specifications for the Embedded PTs and Items 26](#_Toc180396484)

[3.1.2. Embedded PT and Item Format 26](#_Toc180396485)

[3.1.3. Recruitment and Selection of Embedded PT Item Writers 27](#_Toc180396486)

[3.1.4. Embedded PT Item Writer Training 28](#_Toc180396487)

[3.2. Embedded PT and Item Review Process 28](#_Toc180396488)

[3.2.1. Selection of Embedded PTs and Items 28](#_Toc180396489)

[3.2.2. ETS Content Review 29](#_Toc180396490)

[3.2.3. ETS Editorial Review 29](#_Toc180396491)

[3.2.4. ETS Sensitivity and Fairness Review 29](#_Toc180396492)

[3.3. Content Expert Review 30](#_Toc180396493)

[3.3.1. California Educators as Content Experts 30](#_Toc180396494)

[3.3.2. Composition of Item Review Panels 30](#_Toc180396495)

[3.3.3. Meetings for Review of CAA for Science Embedded PTs and Items 31](#_Toc180396496)

[3.4. Data Review Meeting 32](#_Toc180396497)

[Chapter 4: Test Assembly 34](#_Toc180396498)

[4.1. Test Content Specifications and Test Blueprints 34](#_Toc180396499)

[4.1.1. Test Content Specifications 34](#_Toc180396500)

[4.1.2. Test Blueprints 34](#_Toc180396501)

[4.2. Test Design 35](#_Toc180396502)

[4.2.1. Operational Field Test 35](#_Toc180396503)

[4.3. Test Production Process 35](#_Toc180396504)

[4.3.1. Selection of Items 35](#_Toc180396505)

[4.3.2. Content Review of Forms 35](#_Toc180396506)

[4.3.3. CDE Forms Review 35](#_Toc180396507)

[4.3.4. Configuration of the TDS 36](#_Toc180396508)

[References 37](#_Toc180396509)

[Chapter 5: Test Administration 38](#_Toc180396510)

[5.1. Grade Assignment 38](#_Toc180396511)

[5.2. Administration Preparations 38](#_Toc180396512)

[5.2.1. Guides and Videos 38](#_Toc180396513)

[5.2.2. Local Educational Agency (LEA) Training 39](#_Toc180396514)

[5.3. Test Administration 39](#_Toc180396515)

[5.3.1. Administration of the Student Response Check 39](#_Toc180396516)

[5.3.2. Administration of the Embedded PTs 40](#_Toc180396517)

[5.3.3. Administration of the Survey 40](#_Toc180396518)

[5.4. Procedures to Maintain Standardization 40](#_Toc180396519)

[5.4.1. LEA California Assessment of Student Performance and Progress (CAASPP) Coordinator 40](#_Toc180396520)

[5.4.2. CAASPP Test Site Coordinator 41](#_Toc180396521)

[5.4.3. Test Examiners 41](#_Toc180396522)

[5.4.4. Instructions for Test Examiners and Staff Involved in CAA for Science   
Administration 42](#_Toc180396523)

[5.5. Accessibility Features for the Field Test 43](#_Toc180396524)

[5.5.1. Individualizations 43](#_Toc180396525)

[5.5.2. Choice of Administration Scripts 48](#_Toc180396526)

[5.5.3. Choice of Materials 48](#_Toc180396527)

[5.5.4. Type and Level of Accommodations 48](#_Toc180396528)

[5.6. Processing and Scoring 48](#_Toc180396529)

[5.7. Test Security and Confidentiality 48](#_Toc180396530)

[5.7.1. ETS’ Office of Testing Integrity 48](#_Toc180396531)

[5.7.2. Procedures to Maintain Standardization of Test Security 49](#_Toc180396532)

[5.7.3. Security of Electronic Files Using a Firewall 49](#_Toc180396533)

[5.7.4. Transfer of Scores via Secure Data Exchange 50](#_Toc180396534)

[5.7.5. Data Management in the Secure Database 50](#_Toc180396535)

[5.7.6. Statistical Analysis on Secure Servers 50](#_Toc180396536)

[5.7.7. Student Confidentiality 51](#_Toc180396537)

[5.7.8. Security and Test Administration Incident Reporting System (STAIRS) Process 51](#_Toc180396538)

[5.7.9. Appeals 52](#_Toc180396539)

[References 53](#_Toc180396540)

[Chapter 6: Scoring and Reporting 54](#_Toc180396541)

[6.1. CAA for Science Scoring Process 54](#_Toc180396542)

[6.2. Types of Scores 54](#_Toc180396543)

[6.2.1. Percent Correct 54](#_Toc180396544)

[6.2.2. Preliminary Indicator Categories 55](#_Toc180396545)

[6.2.3. Aggregate Score Reporting 59](#_Toc180396546)

[6.3. Survey Questions Regarding Test Administration 60](#_Toc180396547)

[6.3.1. Student Engagement 60](#_Toc180396548)

[6.3.2. Individualization of the Test 60](#_Toc180396549)

[References 62](#_Toc180396550)

[Accessibility Information 63](#_Toc180396551)

[Alternative Text for Equation 6.1 63](#_Toc180396552)

[Appendix 6.A: Demographic Summaries 64](#_Toc180396553)

[Chapter 7: Psychometric Analyses 70](#_Toc180396554)

[7.1. Overview 70](#_Toc180396555)

[7.1.1. Summary of the Analyses 70](#_Toc180396556)

[7.1.2. Sample for the Analyses 71](#_Toc180396557)

[7.2. Classical Item Analysis Statistics 72](#_Toc180396558)

[7.2.1. Classical Item Difficulty Indices (*p*-value and Average Item Score) 72](#_Toc180396559)

[7.2.2. Item Discrimination (Item-Total Correlation) 72](#_Toc180396560)

[7.2.3. Distribution of Item Scores 73](#_Toc180396561)

[7.2.4. Summary of Classical Item Analysis Flagging Criteria 74](#_Toc180396562)

[7.2.5. Classical Item Analysis Results Summary 74](#_Toc180396563)

[7.3. Omission and Completion Rates 75](#_Toc180396564)

[7.3.1. Omit Rates 75](#_Toc180396565)

[7.3.2. Completion Rates 75](#_Toc180396566)

[7.4. Task Difficulty (Overall and by Embedded PT) 76](#_Toc180396567)

[7.5. DIF Analyses 79](#_Toc180396568)

[7.5.1. Dichotomous Items 79](#_Toc180396569)

[7.5.2. Polytomous Items 80](#_Toc180396570)

[7.5.3. DIF Categories and Definitions 81](#_Toc180396571)

[7.5.4. DIF Analysis Results Summary 82](#_Toc180396572)

[7.6. Item Response Theory (IRT) Analyses 83](#_Toc180396573)

[7.6.1. IRT Models 83](#_Toc180396574)

[7.6.2. Item Calibration 83](#_Toc180396575)

[7.7. Reliability Analyses 85](#_Toc180396576)

[7.7.1. Internal Consistency Reliability 86](#_Toc180396577)

[7.7.2. Standard Error of Measurement 87](#_Toc180396578)

[7.8. Validity Evidence 88](#_Toc180396579)

[7.8.1. Evidence in the Design of the CAA for Science 89](#_Toc180396580)

[7.8.2. Evidence Based on Test Content 90](#_Toc180396581)

[7.8.3. Evidence Based on Response Processes 90](#_Toc180396582)

[7.8.4. Evidence Based on Internal Structure 92](#_Toc180396583)

[7.8.5. Evidence Based on Relationship to Other Variables 92](#_Toc180396584)

[References 94](#_Toc180396585)

[Accessibility Information 97](#_Toc180396586)

[Alternative Text for Equation 7.1 97](#_Toc180396587)

[Alternative Text for Equation 7.2 97](#_Toc180396588)

[Alternative Text for Equation 7.3 97](#_Toc180396589)

[Alternative Text for Equation 7.4 97](#_Toc180396590)

[Alternative Text for Equation 7.5 97](#_Toc180396591)

[Alternative Text for Equation 7.6 97](#_Toc180396592)

[Alternative Text for Equation 7.7 97](#_Toc180396593)

[Alternative Text for Equation 7.8 97](#_Toc180396594)

[Alternative Text for Equation 7.9 98](#_Toc180396595)

[Alternative Text for Equation 7.10 98](#_Toc180396596)

[Appendix 7.A: Classical Item Analyses 99](#_Toc180396597)

[Appendix 7.B: Completion Rates 113](#_Toc180396598)

[Appendix 7.C: Distribution of Raw Scores: Total Score Each Embedded PT 115](#_Toc180396599)

[Appendix 7.D: DIF Analysis: Number and Percentage of Items in Each DIF Category 139](#_Toc180396600)

[Appendix 7.E: Item Response Theory (IRT) Analyses Results 143](#_Toc180396601)

[Appendix 7.F: Reliability Estimates 151](#_Toc180396602)

[Appendix 7.G: Validity Analyses 165](#_Toc180396603)

[Chapter 8: Surveys 179](#_Toc180396604)

[8.1. Survey Design and Questionnaire Development 179](#_Toc180396605)

[8.1.1. Student Survey 179](#_Toc180396606)

[8.1.2. Test Examiner Survey 179](#_Toc180396607)

[8.2. Test Examiner Survey Results 179](#_Toc180396608)

[8.2.1. Responses to Background Questions 180](#_Toc180396609)

[8.2.2. Responses Regarding Individualization 181](#_Toc180396610)

[8.2.3. Responses Regarding Helpfulness of the *Administration Planning Guide* 181](#_Toc180396611)

[8.2.4. Responses Regarding the Move to Online Testing 182](#_Toc180396612)

[Appendix 8.A: Distribution of Student Survey Responses 183](#_Toc180396613)

[Survey for Responsive Students 183](#_Toc180396614)

[Chapter 9: Embedded Performance Task and Test Comparability Considerations 185](#_Toc180396615)

[9.1. Considerations 185](#_Toc180396616)

[9.2. Summary of Findings from the Choice of Materials and Level of Individualization Analysis 188](#_Toc180396617)

[9.2.1. Individualization Analysis 188](#_Toc180396618)

[9.2.2. Model Analysis 188](#_Toc180396619)

[9.2.3. Results of the Individualization and Model Analyses 189](#_Toc180396620)

[9.3. Implications for Future Test Administrations 190](#_Toc180396621)

[9.3.1. Key Findings 190](#_Toc180396622)

[9.3.2. Recommendations 191](#_Toc180396623)

[References 192](#_Toc180396624)

[Appendix 9.A: Choice of Materials and Individualization Analysis Data 193](#_Toc180396625)

[Appendix 9.B: Model Analysis Summaries 196](#_Toc180396626)

[Appendix 9.C: Model Analysis Summaries for Intellectual Disabilities Versus Autism 200](#_Toc180396627)

[Appendix 9.D: Model Analysis Summaries for Intellectual Disabilities Versus Learning Disabilities 204](#_Toc180396628)

[Chapter 10: Quality Control Procedures 208](#_Toc180396629)

[10.1. Quality Control of Embedded PT Development 208](#_Toc180396630)

[10.2. Quality Control of Test Assembly and Delivery 208](#_Toc180396631)

[10.2.1. Quality Control of Test Form Development 208](#_Toc180396632)

[10.2.2. Quality Control of Test Assignment 209](#_Toc180396633)

[10.2.3. Quality Control of Test Administration 209](#_Toc180396634)

[10.2.4. Quality Control of Machine-Scoring Procedures 210](#_Toc180396635)

[10.3. Quality Control of Test Materials 210](#_Toc180396636)

[10.3.1. Test Administration Manuals 210](#_Toc180396637)

[10.3.2. Processing Test Materials 211](#_Toc180396638)

[10.4. Quality Control of Psychometric Processes 211](#_Toc180396639)

[10.4.1. Development of Scoring Specifications 211](#_Toc180396640)

[10.4.2. Development of Scoring Procedures 211](#_Toc180396641)

[10.5. Quality Control of Reporting 212](#_Toc180396642)

[References 213](#_Toc180396643)

[Chapter 11: Continuous and Systematic Improvements 214](#_Toc180396644)

[11.1. Improvements from the Second-Year Pilot 214](#_Toc180396645)

[11.1.1. Changes to Test Administration 214](#_Toc180396646)

[11.1.2. Changes to Content 214](#_Toc180396647)

[References 216](#_Toc180396648)

List of Tables

[Acronyms and Initialisms Used in the *California Alternate Assessment for Science Technical Report* x](#_Toc180396649)

[Table 1.1 Organization of the Science Connectors 4](#_Toc180396650)

[Table 2.A.1 CAA for Science Field Test Test-Taking Rates—Registered Students 22](#_Toc180396651)

[Table 2.A.2 CAA for Science Field Test Test-Taking Rates for Test by Student Group, Grade Five 23](#_Toc180396652)

[Table 2.A.3 CAA for Science Field Test Test-Taking Rates for Test by Student Group, Grade Eight 24](#_Toc180396653)

[Table 2.A.4 CAA for Science Field Test Test-Taking Rates for Test by Student Group, High School 25](#_Toc180396654)

[Table 3.1 Number of Items and Points for Each Embedded PT 27](#_Toc180396655)

[Table 3.2 Number of Item Reviewers with Each Qualification 31](#_Toc180396656)

[Table 5.1 Individualizations—Grade Five 45](#_Toc180396657)

[Table 5.2 Individualizations—Grade Eight 46](#_Toc180396658)

[Table 5.3 Individualizations—High School 47](#_Toc180396659)

[Table 5.4 Types of Appeals in CAASPP Testing 52](#_Toc180396660)

[Table 6.1 Indicator Categories 55](#_Toc180396661)

[Table 6.2 Threshold Scores for Preliminary Categories 55](#_Toc180396662)

[Table 6.3 Grade Five Preliminary Indicator Conversion Table 56](#_Toc180396663)

[Table 6.4 Grade Eight Preliminary Indicator Conversion Table 57](#_Toc180396664)

[Table 6.5 High School Preliminary Indicator Conversion Table 58](#_Toc180396665)

[Table 6.6 Demographic Student Groups to Be Reported 59](#_Toc180396666)

[Table 6.A.1 Demographic Summary for Grade Five 64](#_Toc180396667)

[Table 6.A.2 Demographic Summary for Grade Eight 66](#_Toc180396668)

[Table 6.A.3 Demographic Summary for High School 68](#_Toc180396669)

[Table 7.1. Analysis Data Sources 71](#_Toc180396670)

[Table 7.2 Classical Item Statistics 74](#_Toc180396671)

[Table 7.3 Raw Score Summary for Each Embedded PT 77](#_Toc180396672)

[Table 7.4 DIF Categories for Dichotomous Items 81](#_Toc180396673)

[Table 7.5 DIF Categories for Polytomous Items 81](#_Toc180396674)

[Table 7.6 Student Groups for DIF Comparison 82](#_Toc180396675)

[Table 7.7 Summary Statistics for Raw Scores and Stratified Alpha 87](#_Toc180396676)

[Table 7.8 Correlations Among the CAAs for ELA, Mathematics, and Science 93](#_Toc180396677)

[Table 7.A.1 Classical Item Statistics for Each Embedded PT 100](#_Toc180396678)

[Table 7.A.2 Average Item Score and Polyserial Correlation for Each Item, Grade Five 101](#_Toc180396679)

[Table 7.A.3 Average Item Score and Polyserial Correlation for Each Item, Grade Eight 103](#_Toc180396680)

[Table 7.A.4 Average Item Score and Polyserial Correlation for Each Item, High School 105](#_Toc180396681)

[Table 7.A.5 Distribution of Item Scores for Each Item, Grade Five 107](#_Toc180396682)

[Table 7.A.6 Distribution of Item Scores for Each Item, Grade Eight 109](#_Toc180396683)

[Table 7.A.7 Distribution of Item Scores for Each Item, High School 111](#_Toc180396684)

[Table 7.B.1 Percentage of Students in Each Grade Completing Embedded PTs 113](#_Toc180396685)

[Table 7.B.2 Completion Rates by Grade Level for Each Embedded Performance Task   
(PT) 114](#_Toc180396686)

[Table 7.C.1 Distribution of Total Score and PT Scores, Grade Five Version One 115](#_Toc180396687)

[Table 7.C.2 Distribution of Total Score and PT Scores, Grade Five Version Two 117](#_Toc180396688)

[Table 7.C.3 Distribution of Total Score and PT Scores, Grade Eight Version One 119](#_Toc180396689)

[Table 7.C.4 Distribution of Total Score and PT Scores, Grade Eight Version Two 121](#_Toc180396690)

[Table 7.C.5 Distribution of Total Score and PT Scores, High School Version One 123](#_Toc180396691)

[Table 7.C.6 Distribution of Total Score and PT Scores, High School Version Two 125](#_Toc180396692)

[Table 7.C.7 Distribution of Total Score and PT Scores, Grade Ten Version One 127](#_Toc180396693)

[Table 7.C.8 Distribution of Total Score and PT Scores, Grade Ten Version Two 129](#_Toc180396694)

[Table 7.C.9 Distribution of Total Score and PT Scores, Grade Eleven Version One 131](#_Toc180396695)

[Table 7.C.10 Distribution of Total Score and PT Scores, Grade Eleven Version Two 133](#_Toc180396696)

[Table 7.C.11 Distribution of Total Score and PT Scores, Grade Twelve Version One 135](#_Toc180396697)

[Table 7.C.12 Distribution of Total Score and PT Scores, Grade Twelve Version Two 137](#_Toc180396698)

[Table 7.D.1 Categorization of DIF, Grade Five 139](#_Toc180396699)

[Table 7.D.1 Categorization of DIF, Grade Five (Continued) 139](#_Toc180396700)

[Table 7.D.2 Categorization of DIF, Grade Eight 140](#_Toc180396701)

[Table 7.D.2 Categorization of DIF, Grade Eight (Continued) 140](#_Toc180396702)

[Table 7.D.3 Categorization of DIF, High School 141](#_Toc180396703)

[Table 7.D.3 Categorization of DIF, High School (Continued One) 141](#_Toc180396704)

[Table 7.D.3 Categorization of DIF, High School (Continued Two) 142](#_Toc180396705)

[Table 7.D.4 Items Exhibiting Significant DIF, Grade Five 142](#_Toc180396706)

[Table 7.E.1 IRT Parameter Estimates for All CAA for Science Items 143](#_Toc180396707)

[Table 7.E.2 IRT Item Difficulty, Grade Five 143](#_Toc180396708)

[Table 7.E.3 IRT Item Difficulty, Grade Eight 145](#_Toc180396709)

[Table 7.E.4 IRT Item Difficulty, High School 146](#_Toc180396710)

[Table 7.E.5 Item Difficulties and Omit Rate, Grade Five 147](#_Toc180396711)

[Table 7.E.6 Item Difficulties and Omit Rate, Grade Eight 149](#_Toc180396712)

[Table 7.E.7 Item Difficulties and Omit Rate, High School 150](#_Toc180396713)

[Table 7.F.1 Reliabilities and Standard Errors of Measurement (SEMs) by Gender 151](#_Toc180396714)

[Table 7.F.2 Reliabilities and SEMs by Ethnicity 152](#_Toc180396715)

[Table 7.F.2 Reliabilities and SEMS by Ethnicity (Continued One) 153](#_Toc180396716)

[Table 7.F.2 Reliabilities and SEMS by Ethnicity (Continued Two) 154](#_Toc180396717)

[Table 7.F.3 Reliabilities and SEMs by English Proficiency 155](#_Toc180396718)

[Table 7.F.3 Reliabilities and SEMs by English Proficiency (Continued One) 156](#_Toc180396719)

[Table 7.F.3 Reliabilities and SEMs by English Proficiency (Continued Two) 157](#_Toc180396720)

[Table 7.F.4 Reliabilities and SEMs by Economic Status 158](#_Toc180396721)

[Table 7.F.5 Reliabilities and SEMs by Migrant Status 159](#_Toc180396722)

[Table 7.F.6 Reliabilities and SEMs by Primary Disabilities 160](#_Toc180396723)

[Table 7.F.6 Reliabilities and SEMs by Primary Disabilities (Continued One) 161](#_Toc180396724)

[Table 7.F.6 Reliabilities and SEMs by Primary Disabilities (Continued Two) 162](#_Toc180396725)

[Table 7.F.6 Reliabilities and SEMs by Primary Disabilities (Continued Three) 163](#_Toc180396726)

[Table 7.F.6 Reliabilities and SEMs by Primary Disabilities (Continued Four) 164](#_Toc180396727)

[Table 7.G.1 Correlations Between Raw Scores and Test Engagement Response 165](#_Toc180396728)

[Table 7.G.2 Raw Score by PT Engagement Response, Grade Five 166](#_Toc180396729)

[Table 7.G.3 Raw Score by PT Engagement Response, Grade Eight 167](#_Toc180396730)

[Table 7.G.4 Raw Score by PT Engagement Response, High School 168](#_Toc180396731)

[Table 7.G.5 Raw Score by PT Engagement Response, Grade Ten 169](#_Toc180396732)

[Table 7.G.6 Raw Score by PT Engagement Response, Grade Eleven 170](#_Toc180396733)

[Table 7.G.7 Raw Score by PT Engagement Response, Grade Twelve 171](#_Toc180396734)

[Table 7.G.8 Total Testing Time (In Minutes) by Grade and Version 172](#_Toc180396735)

[Table 7.G.9 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Five 173](#_Toc180396736)

[Table 7.G.10 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Eight 174](#_Toc180396737)

[Table 7.G.11 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, High School 175](#_Toc180396738)

[Table 7.G.12 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Ten 176](#_Toc180396739)

[Table 7.G.13 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Eleven 177](#_Toc180396740)

[Table 7.G.14 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Twelve 178](#_Toc180396741)

[Table 8.1 Teacher for Student Tested 180](#_Toc180396742)

[Table 8.2 How Many Students Tested Per Test Examiner 180](#_Toc180396743)

[Table 8.3 Grade Administered 181](#_Toc180396744)

[Table 8.4 Frequency of Individualization 181](#_Toc180396745)

[Table 8.5 Helpfulness of the *Administration Planning Guide* 181](#_Toc180396746)

[Table 8.6 Preference for Online Format 182](#_Toc180396747)

[Table 8.A.1 Responses to Question 1, “How engaged was the student with this performance task?” 183](#_Toc180396748)

[Table 8.A.2 Responses to Question 2, “Did you individualize any aspect of Orienting Activity #1 and the first five test questions, where permitted?” 183](#_Toc180396749)

[Table 8.A.3 Responses to Question 3, “Did you individualize any aspect of Orienting Activity #2 and the first five test questions, where permitted?” 184](#_Toc180396750)

[Table 9.1 Linear Models Estimated in Material Choices Analysis 187](#_Toc180396751)

[Table 9.2 Regroupings of the Disability Types 187](#_Toc180396752)

[Table 9.A.1 Individualizations—Grade Five 193](#_Toc180396753)

[Table 9.A.2 Individualizations—Grade Eight 194](#_Toc180396754)

[Table 9.A.3 Individualizations—High School 195](#_Toc180396755)

[Table 9.B.1 Model Summary—Grade Five, Physical Sciences, Activity 2 196](#_Toc180396756)

[Table 9.B.2 Model Summary—Grade Eight, Physical Sciences, Activity 1 196](#_Toc180396757)

[Table 9.B.3 Model Summary—Grade Eight, Physical Sciences, Activity 2 197](#_Toc180396758)

[Table 9.B.4 Model Summary—High School, Life Sciences, Activity 1 197](#_Toc180396759)

[Table 9.B.5 Model Summary—High School, Physical Sciences, Activity 1 198](#_Toc180396760)

[Table 9.B.6 Model Summary—High School, Physical Sciences, Activity 2 198](#_Toc180396761)

[Table 9.B.7 Model Summary—High School, Earth and Space Sciences, Activity 1 199](#_Toc180396762)

[Table 9.C.1 Model Summary—Grade Five, Physical Sciences, Activity 2 200](#_Toc180396763)

[Table 9.C.2 Model Summary—Grade Eight, Physical Sciences, Activity 1 200](#_Toc180396764)

[Table 9.C.3 Model Summary—Grade Eight, Physical Sciences, Activity 2 201](#_Toc180396765)

[Table 9.C.4 Model Summary—High School, Life Sciences, Activity 1 201](#_Toc180396766)

[Table 9.C.5 Model Summary—High School, Physical Sciences, Activity 1 202](#_Toc180396767)

[Table 9.C.6 Model Summary—High School, Physical Sciences, Activity 2 202](#_Toc180396768)

[Table 9.C.7 Model Summary—High School, Earth and Space Sciences, Activity 1 203](#_Toc180396769)

[Table 9.D.1 Model Summary—Grade Five, Physical Sciences, Activity 2 204](#_Toc180396770)

[Table 9.D.2 Model Summary—Grade Eight, Physical Sciences, Activity 1 204](#_Toc180396771)

[Table 9.D.3 Model Summary—Grade Eight, Physical Sciences, Activity 2 205](#_Toc180396772)

[Table 9.D.4 Model Summary—High School, Life Sciences, Activity 1 205](#_Toc180396773)

[Table 9.D.5 Model Summary—High School, Physical Sciences, Activity 1 206](#_Toc180396774)

[Table 9.D.6 Model Summary—High School, Physical Sciences, Activity 2 206](#_Toc180396775)

[Table 9.D.7 Model Summary—High School, Earth and Space Sciences, Activity 1 207](#_Toc180396776)

Acronyms and Initialisms Used in the *California Alternate Assessment for Science Technical Report*

|  |  |
| --- | --- |
| Term | Definition |
| 1PL | one parameter logistic |
| AERA | American Educational Research Association |
| AI | artificial intelligence |
| AIR | American Institutes for Research |
| AIS | average item score |
| APA | American Psychological Association |
| CA NGSS | California Next Generation Science Standards |
| CAA | California Alternate Assessment |
| CAASPP | California Assessment of Student Performance and Progress |
| CALPADS | California Longitudinal Pupil Achievement Data System |
| CalTAC | California Technical Assistance Center |
| CAST | California Science Test |
| *CCR* | *California Code of Regulations* |
| CCSS | Common Core State Standards |
| CDE | California Department of Education |
| CERS | California Educator Reporting System |
| CSA | California Spanish Assessment |
| *DFA* | *Directions for Administration* |
| DIF | differential item functioning |
| *EC* | *Education Code* |
| ELA | English language arts/literacy |
| ELD | English Language Development |
| eSKM | Enterprise Score Key Management |
| ETS | Educational Testing Service |
| EUs | essential understandings |
| FKSA | focal knowledge, skills, and abilities |
| GPCM | generalized partial credit model |
| IEP | individualized education program |
| IRT | item response theory |
| ISAAP | Individual Student Assessment Accessibility Profile |
| LEA | local educational agency |
| MH DIF | Mantel-Haenszel differential item functioning |
| MI | Measurement Incorporated |
| NCME | National Council on Measurement in Education |
| ORS | Online Reporting System |
| OTI | Office of Testing Integrity |
| PCM | partial credit model |
| PE | performance expectation |
| PT | performance task |
| QA | quality assurance |
| QC | quality control |
| RMSEA | root mean square error of approximation |
| SBE | State Board of Education |
| Science Connectors | Science Core Content Connectors |
| SD | standard deviation |
| SEM | standard error of measurement |
| SFTP | Secure File Transfer Protocol |
| SMD | standardized mean difference |
| SR | selected response |
| SRC | Student Response Check |
| SS3 | type III sum of squares |
| STAIRS | Security and Test Administration Incident Reporting System |
| TDS | test delivery system |
| TOMS | Test Operations Management System |
| UAT | user acceptance testing |
| USC | United States Code |

**This page is intentionally left blank.**

## Introduction

### Background

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

California adopted the California Next Generation Science Standards (CA NGSS) in September 2013. The California Alternate Assessment (CAA) for Science is an assessment aligned with the Science Core Content Connectors (Science Connectors) derived from the CA NGSS. Its field test was administered during the 2018–2019 CAASPP administration.

The CAA for Science is designed for students with the most significant cognitive disabilities and measures what students know and can do in science. The purposes of the CAA for Science are to measure what students know and can do based on the Science Connectors linked to the CA NGSS across the three science domains and help identify and address gaps in knowledge or skills early so students can receive the support they need (California Department of Education [CDE], 2020a).

The CAA for Science is for students in grades five and eight and in high school whose individualized education program (IEP) teams have determined that alternate assessments are appropriate for the student (California Department of Education [CDE], 2019). Note that this technical report focuses on the CAA for Science and *not* the CAAs for English language arts/literacy (ELA) and mathematics, which are reported upon separately.

In 2018–2019, the CAASPP System comprised the following assessments:

* Smarter Balanced assessments and tools for the general student population:
* Summative Assessments—Online assessments for 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 (CCSS) that may be administered to students at any grade level
* Digital Library—Professional development materials and instructional resources designed to help teachers use formative assessment processes for improved teaching and learning in all grades
* CAAs for ELA and mathematics in grades three through eight and grade eleven for students with significant cognitive disabilities
* Science assessments in grades five and eight and high school (grade ten, eleven, or twelve; these are the California Science Test [CAST] and the CAA for Science)
* The California Spanish Assessment (CSA), optional for eligible students in grades three through eight and high school and designed to measure a student’s Spanish competency in reading, writing mechanics, and listening, as well as a high school measure suitable to be used in part for the California Seal of Biliteracy

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

### Purpose of the Field Test

The purpose of the CAA for Science field test was to use the previously piloted, embedded performance task (PT) format selected for use in assessing the Science Connectors derived from the CA NGSS for the CAA-eligible student population. The Science Connectors provide learning goals that are aligned appropriately with the needs of students with the most significant cognitive disabilities and serve as the basis for the state’s CA NGSS alternate summative science assessments for eligible students. This field test was delivered entirely online. The overarching goal of the field test was to build on what was learned from the first and second pilot tests, laying the groundwork for developing a final blueprint and test design for future field testing and, ultimately, the launch of the CAA for Science operational assessments.

The transition from a paper-based administration to delivery online introduced change as a function in the difference in the format. However, it was the feedback and results of the pilot observational studies as well as what was indicated in the results of the test examiner survey after the paper-based pilot that were used to inform significant adjustments in the design of the online field test. For example, feedback from the observations and survey indicated that test examiners had difficulty in handling various test materials and needed additional time to prepare for orienting activities. In response to this feedback, the following major changes were introduced for the online field test:

* **Simplified orienting activities:** The orienting activities were significantly simplified; additionally, the activities chosen were ones that would require only materials readily available in the classroom.
* **Video option:** Test examiners were permitted to show a video within the test delivery system (TDS) in lieu of a hands-on orienting activity, with the hands-on activity being offered as an option for individualization.
* **More specificity in individualization:** More specific instruction was given as to which elements of the test could and could not be individualized.

These changes to the orienting activities and to the items reduced the amount of materials and preparation time required of the test examiners. The current design includes only a very simple activity to orient the student to a key concept being assessed; no subsequent questions are based on the student’s experience with that orienting activity. Some items still incorporate very simple activities, such as rolling a ball or pushing an object.

### Field Test Content and Design

#### Assessment Model

The California State Board of Education (SBE) approved the conceptual design for the CAA for Science in July 2016. This design uses an embedded PT design, meaning that each embedded PT is expected to be administered shortly after content related to the Science Connectors has been taught. Test examiners administer a set of test questions measuring two Science Connectors from one of the three science domains (CDE, 2018).

In cases where implementation has been particularly successful, alternate assessments based on a collection of embedded PTs (sometimes referred to as a “body of evidence”)have been shown to leverage higher academic learning expectations for students taking an alternate assessment while promoting enhanced curricular and instructional supports for teachers (Gong & Marion, 2006).

The guiding principles adopted for the CAA for Science are that these assessments

* support and promote teachers’ implementation of the CA NGSS;
* embed summative assessment into instructional practice;
* offer a developmentally appropriate opportunity for students with the most significant cognitive disabilities to be assessed on their science knowledge, skills, and abilities; and
* provide meaningful information about academic performance to both parents/‌guardians and teachers.

California’s relatively small population of students with the most significant cognitive disabilities who are eligible for an alternate science assessment[[1]](#footnote-2) also makes the use of this assessment model reasonable.

#### California Next Generation Science Standards Core Content Connectors (Science Connectors)

The assessment is aligned with the Science Connectors. The Science Connectors are the appropriate standards for the student population assigned to take the CAA for Science. The Science Connectors bridge the CA NGSS performance expectations (PEs) for the standard student population to the expectations developed to provide appropriate levels of challenge and rigor for students with the most significant cognitive disabilities. Table 1.1 summarizes the structure and organization of the Science Connectors.

Table 1.1 Organization of the Science Connectors

|  |  |
| --- | --- |
| Assessment Components | Grade Level (Kindergarten–12) |
| Performance Expectation | Incorporates a disciplinary core idea, a science and engineering practice, and a crosscutting concept into an assessable statement of what students should know and be able to accomplish with regard to the four domains (i.e., Life Sciences; Physical Sciences; Earth and Space Sciences; and Engineering, Technology, and Applications of Science) |
| Science Connector | Builds a bridge to the content of a CA NGSS PE |
| Focal Knowledge, Skills, and Abilities (FKSA) | Describes what students should know and be able to do in terms of the Science Connector (FKSA1 up to FKSA6) |
| Essential Understanding | Defines a basic, foundational key idea or concept |

#### Test Components for the Field Test

The 2018–2019 field test of the CAA for Science involved four components:

1. Three embedded PTs
2. A brief student survey
3. An optional test examiner survey
4. An optional training embedded PT

##### Embedded Performance Tasks

An embedded PT represents the model of assessment known as curriculum-embedded PTs. The intent behind this assessment model is to have educators embedding PTs as summative assessments following classroom instructional activities relating to the Science Connectors.

For the 2018–2019 CAASPP administration, embedded PTs per grade were tested for the CAA for Science: three for grade five, three for grade eight, and three for high school (i.e., grade ten, eleven, or twelve). Each embedded PT included information for the test examiner, describing the hands-on activity and how to administer the embedded PT items. The embedded PT item types included selected-response, match, and grid items; these are described in subsection [*3.1.2 Embedded PT and Item Format*](#_Embedded_PT_and).

The secure embedded PTs were delivered to students through the CAASPP test delivery system (TDS), and the *Directions for Administration* (*DFAs*) were delivered to LEAs as downloadable PDFs within the Test Operations Management System (TOMS). Test examiners administered the embedded PTs in one-on-one sessions with the answers recorded in the TDS.

##### Survey for Students

During the 2018–2019 administration year, students were asked to respond to a survey administered by test examiners. After the task was administered to the student, test examiners would then solicit student responses to a short survey. The purpose of the student survey was to collect basic information about students’ experiences with the assessment process.

This survey was included in the last section of the embedded PT delivered through the TDS. The survey had four questions about any individualizations used during the test administration and one question on the level of student engagement. Refer to [chapter 8](#_Surveys) for additional information about the student survey design and results.

##### Optional Test Examiner Survey

An optional survey was presented to test examiners to obtain teachers’ feedback on the field test administration and assessment processes in order to guide the implementation of each respective assessment. This survey was linked on the CAASPP Portal and hosted on SurveyGizmo.com, a website with survey-creation and hosting services. Refer to [chapter 8](#_Surveys) for additional information about the optional test examiner survey and results.

##### Optional Training Embedded PT

Test examiners had an opportunity to gain familiarity with the new assessment and embedded PT format through a training embedded PT made available on caaspp.org and the TDS.

### Intended Population

All eligible students enrolled in grades five, eight, and high school whose IEP indicated an alternate assessment were selected by the LEA to take the CAA for Science (*California Code of Regulations*, Title 5 [5*CCR*]Education, Division 1, Chapter 2, Subchapter 3.75, Article 1, Section 851.5[c]). High school students in an ungraded program whose calculated grade was twelve might also have taken this assessment, as did students in grades ten or eleven, if selected by the LEA to test.

For students with significant cognitive disabilities, the decision to administer the CAST or the CAA for Science was made by their IEP team. Parents/Guardians may submit a written request to have their child opted out from taking any or all parts of the CAAs. Students whose parents/guardians submit a written request may opt out of taking the tests (*Education Code [EC]* Section 60615).

### Testing Window and Times

For the 2018–2019 CAASPP administration, the CAA for Science field test embedded PTs were available for administration on or after January 8, 2019, through the last day of instruction at the LEA or July 15, 2019, whichever came first (5 *CCR,* Section 855[a][2]).

Similar to other CAASPP assessments, the CAA for Science field test embedded PTs were untimed for test takers. This assessment was administered individually, and testing time varied from one student to another on the basis of factors such as the student’s response time and attention span. Administration of the CAA for Science field test embedded PTs occurred over as many days as required to meet a student’s needs.

### Significant Developments for the CAA for Science 2018–2019 Field Test

#### Online Delivery

The CAA for Science was administered as an online assessment in the 2018–2019 administration. Student responses were entered directly into the TDS.

### Limitations of the Field Test

The CAA for Science field test aligned with, and measured against, the CA NGSS Science Connectors. Development was challenging because of the distinct difference between the new and previous California science standards. Because the purpose of the field test was to evaluate the items rather than students’ knowledge, skills, and abilities, the field test was not a full representation of the assessment model for the CA NGSS Science Connectors.

### Groups and Organizations Involved with the Assessment

#### State Board of Education (SBE)

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

In addition to adopting the rules and regulations for itself, its appointees, and California’s public schools, the SBE also is the state educational agency responsible for overseeing California’s compliance with programs that meets the requirements of the federal Every Student Succeeds Act and the state’s Public School Accountability Act, which measures the academic performance and progress of schools on a variety of academic metrics (CDE, 2020b).

#### California Department of Education (CDE)

The CDE oversees California’s public school system, which is responsible for the education of more than 6,200,000 children and young adults in more than 10,500 schools.[[2]](#footnote-3) 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, prepares students to live, work, and thrive in a highly connected world.

Within the CDE, it is the Instruction & Measurement branch that oversees programs promoting improved student achievement. Programs include oversight of statewide assessments and the collection and reporting of educational data (CDE, 2020c).

#### California Educators

A variety of California educators, including school administrators and teachers experienced in teaching students with cognitive disabilities—who were selected based on their qualifications, experiences, demographics, and geographic locations—were invited to participate in the entire CAA for Science assessment development process. This included defining the purpose and scope of the assessment, assessment design, item development, data review, and score reporting.

#### Contractors

##### Educational Testing Service (ETS)

The CDE and the SBE contract with ETS to develop and administer the CAA for Science. As the prime contractor, ETS has the overall responsibility for working with the CDE to implement and maintain an effective assessment system and to coordinate the work of its subcontractors. Activities directly conducted by ETS include but are not limited to the following:

* Providing management of the program activities
* Supporting and training counties, LEAs, and direct funded charter schools
* Providing tiered help-desk support to LEAs
* Hosting and maintaining a website with resources for LEA CAASPP coordinators
* Developing, hosting, and providing support for Test Operations Management System (TOMS)
* Developing all CAA for Science embedded PTs
* Constructing, producing, and controlling the quality of CAASPP test forms and related test materials, including grade- and content-specific *DFAs*
* Processing student test assignments
* Completing all psychometric procedures
* Producing and distributing score reports
* Developing a score reporting website that can be viewed by the public

##### American Institutes for Research (AIR)

ETS also monitors and manages the work of AIR (now Cambium Assessment), subcontractor to ETS for the CAASPP System of online assessments. Activities conducted by AIR include

* providing the AIR proprietary TDS, including the Student Testing Interface, Test Administrator Interface, secure browser, and training tests;
* hosting and providing support for its TDS and the Online Reporting System (ORS), a component of the overall CAASPP Assessment Delivery System;
* scoring machine-scorable items; and
* providing Level 3 technology help desk support to LEAs.

### Systems Overview and Functionality

#### Test Operations Management System (TOMS)

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

* Managing test administration windows
* Assigning CAA test examiner user roles
* Managing student test assignments and accessibility resources
* Viewing and downloading reports
* Providing a platform for authorized user access to secure materials such as CAA for Science *DFAs*, CAASPP user information, and access to the *CAASPP Security and Test Administration Incident Reporting System* form

TOMS receives student enrollment data and LEA and school hierarchy data from the California Longitudinal Pupil Achievement Data System (CALPADS) via a 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.”[[3]](#footnote-4) LEA staff involved in the administration of the CAA for Science—such as LEA CAASPP coordinators, CAASPP test site coordinators, 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 examiner cannot download student reports. A description of user roles is explained more extensively in the *2018–19 CAASPP Online Test Administration Manual* (CDE, 2019).

#### Test Delivery System (TDS)

The CAASPP TDS is the means by which the statewide online assessments are delivered to students. Components of the TDS include

* the Test Administrator Interface, the web browser–based application that allows test administrators and test examiners to activate student tests and monitor student testing;
* the Student Testing Interface, on which students take the CAA for Science using the secure browser and with assistance from the test examiner as needed; and
* the secure browser, the online application through which the Student Testing Interface may be accessed. The secure browser prevents students from accessing other applications during testing.

#### Training Test

The publicly available training test is provided to prepare students for the summative assessment. This test, available for grades five and eight and high school, simulates the experience of the CAA for Science online assessments. Training tests align with PEs but do not produce scores. Students may access them using a web browser.

The purposes of the training test are to

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

#### Online Reporting System (ORS) and California Educator Reporting System (CERS)

Currently, there are two California online reporting systems: the ORS and the CERS. Over the next two years, CERS will become the single resource where LEA staff accesses student results from the summative and interim CAASPP assessments as well as results from the English Language Proficiency Assessments for California.

The ORS is the system used by LEAs to view preliminary student results from the CAASPP assessments. The primary purposes of the ORS are for LEAs to access completion data to determine which students need to complete testing or start testing, and for LEAs to access preliminary score reports that can provide claim-related data for schools within the LEA. Results in the ORS are preliminary and may not be used for accountability purposes.

The CERS allows educators to view their students’ assessment results using grouping and other new features. For example, educators can create customized groups from assigned student groups; for interim assessments, specific assessment items can be viewed with student responses; and a distractor analysis feature can be used to identify student strengths and needs.

### Overview of the Technical Report

This technical report addresses the characteristics of the CAAs for Science administered starting in January 2019 and contains 10 additional chapters as follows:

* [Chapter 2](#_Overview_of_the) presents an overview of processes involved in the CAA for Science field test, including descriptions of item development, test administration, and psychometric analyses.
* [Chapter 3](#_Embedded_PT_Development) discusses the detailed procedures of embedded PT development for the CAA for Science field test.
* [Chapter 4](#_Test_Assembly) describes the process of test assembly for the CAA for Science field test.
* [Chapter 5](#_Test_Administration) describes the details of administering the embedded PTs for the CAA for Science field test, as well as the procedures followed by ETS to ensure test security.
* [Chapter 6](#_Scoring_and_Reporting_1) summarizes the scoring approaches and type of scores that are reported for the CAA for Science field test.
* [Chapter 7](#_Psychometric_Analyses) summarizes the statistical procedures and results for 2018–2019, including classical item analyses, test completion rates and analyses, and differential item functioning analyses.
* [Chapter 8](#_Surveys) describes the development and administration of the survey questionnaires for test examiners and the results of analyses on their responses.
* [Chapter 9](#_Toc30746583) presents the results of an investigation conducted to evaluate the impact of both the choice of materials and choice to individualize on the performance of the embedded PTs administered as part of the 2018–2019 CAA for Science field test.
* [Chapter 10](#_Quality_Control_Procedures) discusses the various procedures used to ensure the quality of the CAA for Science field test.
* [Chapter 11](#_Continuous_and_Systematic) discusses the various procedures used to gather information to improve the CAA for Science as well as strategies to implement possible improvements.

### References

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

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacrament, CA: California Department of Education.

California Department of Education. (2019). *CAASPP online test administration manual, 2018–19 administration.* Sacramento, CA: California Department of Education.

California Department of Education. (2020a). *California Alternate Assessment for Science*. Sacramento, CA: California Department of Education.

California Department of Education. (2020c). *Organization.*

California Department of Education. (2020b). *State Board of Education responsibilities.*

Gong, B., & Marion, S. (2006). Dealing with flexibility in assessments for students with significant cognitive disabilities. *Synthesis Report, 60*.

## Overview of the CAA for Science Field Test Processes

This chapter provides an overview of the processes implemented by Educational Testing Service (ETS) during the full testing cycle for the 2018–2019 California Alternate Assessment (CAA) for Science, including descriptions of item development, test administration, accessibility resources, and psychometric analyses.

### Embedded Performance Task (PT) and Item Development and Review

As part of the adaptation and alignment process, ETS developed all embedded PTs for the CAA for Science in accordance with the *ETS* *Standards for Quality and Fairness* (2014).

#### Selection of Science Connectors for Embedded PT Development

For the field test, ETS developed three embedded PTs for each grade or grade band according to the blueprint (California Department of Education [CDE], 2018a). The State Board of Education (SBE)–approved blueprint document identifies the California Next Generation Science Standards (CA NGSS) Core Content Connectors (Science Connectors) eligible to be assessed through embedded PTs. The blueprint was developed in consultation with the CDE. It consists of a Science Connector prioritization plan based on input from California educators, other internal and external experts on both the CA NGSS and alternate assessments. Each of the embedded PTs assesses two of these Science Connectors.

#### Embedded PT Development for Grades Five and Eight and High School

ETS developed each embedded PT as a set of items assessing a particular Science Connector. Each set of items was associated with a particular concept or phenomenon. The concept or topic selected for each embedded PT was reviewed to ensure that the content and presentation were accessible to, and developmentally appropriate for, students with the most significant cognitive disabilities.

A full review of the process to develop embedded PTs, including the number of items and the type of items, can be found in [chapter 3](#_Embedded_PT_Development).

##### Task Format

The CAA for Science includes the following primary online item formats:

* **Selected response (SR) items—**Students are instructed to select one or more choices. Most CAA items have two or three options; a few items have four options.
* **Technology-enhanced items**—Technology beyond simple option selection is incorporated in some items. These items can resemble simple classroom activities in which students might complete a diagram or make a selection from information in a chart.

Detailed information on item format is included in subsection [*3.1.2 Embedded PT and Item Format*](#_Embedded_PT_and) in [*Chapter 3: Embedded PT Development and Review*](#_Embedded_PT_Development).

SR and technology-enhanced items have either one or two points and are machine-scored.

##### Item Specifications

The CAA item specifications provide descriptions of item characteristics that are intended to measure each content standard consistently. They were developed based on the CA NGSS guidelines and clarifications from the Science Connectors and essential understandings (EUs). During item development, item developers were provided CAA item specifications and a CAA style guide that contained detailed information about the consistency in item development and item review processes. Refer to subsection [*3.1.1 Specifications for the Embedded PTs and Items*](#_Specifications_for_the) in [chapter 3](#_Embedded_PT_Development) for detailed information about item specifications.

##### Item Banking

The test forms of the 2018–2019 CAA for Science administration were comprised of newly developed, embedded field test items.

After the 2018–2019 CAA administration, initial item analyses were implemented, and the results were reviewed by ETS psychometric and assessment 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 subsection [*7.2 Classical Item Analysis Statistics*](#_Classical_Item_Analysis).

Next, the field test items were calibrated. Refer to subsection [*7.6.2 Item Calibration*](#_Item_Calibration) for a description of this process.

Content experts from ETS and the CDE, as well as selected California educators, reviewed the associated item statistics and evaluated the performance of items during the annual data review meeting. They also reviewed the flagged items—those whose statistics fall beyond expected ranges—and worked to provide plausible explanations for these particular items based on their knowledge of the student population.

With the CDE’s approval, the field test items, together with their statistical information, were entered into the item bank for form assembly in future administrations. It is expected that more new items will be developed, field-tested, and entered into the item bank after the 2018–2019 administration. In this way, the item bank will expand gradually to support future operational forms.

#### Universal Design Principles

The application of universal design in assessment development involves establishing that tests and testing environments are usable by all students to the greatest extent possible. To allow for the widest possible range of students taking the CAA for Science, ETS trains all item writers to follow the principles of universal design in their development and revision of test items. These principles include, but are not limited to

* reducing wordiness;
* avoiding complex sentence structures and sentences that begin with dependent clauses;
* avoiding ambiguity;
* breaking up compound sentences;
* avoiding colloquialisms and words with double meanings;
* using active tense when possible;
* selecting developmentally appropriate text levels and terminology; and
* consistently applying concept names and graphic conventions.

Universal design principles also inform decisions about test layout and design, including such features as type size, line length, spacing, and graphics. These principles provide flexibility for the ways information is presented as well as for the ways students are engaged with, and respond to, that information. The goal is to reduce barriers in assessing *all* students.

### Test Assembly

The 2018–2019 field test was assembled in accordance with the CAA for Science blueprint, which was approved by the SBE in January 2018 (CDE, 2018a). The CAA for Science is a linear form comprised of three embedded PTs, each comprised of two Connector sets that assess Science Connectors from one of the three science domains.

The assembly began with selection of approved field test–ready items from the item bank. After the initial assembly, test developers reviewed the assembled forms using comprehensive checklists to evaluate blueprint alignment, item content, clueing and content overlap, and overall balance of content with regard to gender and ethnicity representation, variety of item types, and so forth.

After test developers assembled and reviewed the draft test forms, the forms were submitted for psychometric review and approval. Approved forms then received additional content and editorial reviews, including key checks and review of scoring files, before being submitted to the CDE for review and feedback. After responding to feedback from the CDE, forms received a final content review to ensure any requested revisions were administered accurately before submittal to the CDE for their approval.

#### Test Design

The CAA for Science is based on a linear design comprised of three embedded PTs, each comprised of two Connector sets that assess standards from one of the three science domains of Life Sciences, Physical Sciences, and Earth and Space Sciences. The Connector sets also incorporate contexts aligned to the Engineering, Technology, and Application of Science domain.

Connector sets are groups of five items, along with an orienting activity, that assess a Science Connector. Two Connector sets are paired to create an embedded PT that consists of 10 items and two orienting activities.

The three embedded PTs were intended to be administered throughout the school year, immediately after students received instruction in the Science Connectors assessed by the embedded PT. Thus, the embedded PTs could be administered in any order throughout the instructional year.

The operational field test embedded PTs were available for administration from January 8, 2019, through July 15, 2019. In subsequent operational years, the embedded PTs will be available for administration beginning in September of each year.

#### Test Blueprints

Test blueprints specify the total number of items on each test and the number of items in each content category according to standards (CDE, 2018a). The standards upon which CAA for Science test blueprints are built consist of the Science Connectors and EUs, both derived from the CCSS. The blueprints for the CAA for Science were adopted by the SBE in January 2018.

The CAA for Science test blueprints are unique to each grade level and content area. These blueprints designate the breakdown first by content category and then by Science Connectors. Information on each grade-level test blueprint includes

* specific ratio of each content category or domain on the overall test,
* specific Science Connectors to be assessed, and
* the maximum number of total items.

#### Test Length

The number of items in each of the CAAs for Science is the same across grades—there are 10 items per embedded PT. For the field test, each student was given three embedded PTs.

Refer to subsection [*4.2 Test Design*](#_Test_Design) in [*Chapter 4: Test Assembly*](#_Test_Assembly) for more details on test form assembly.

### Test Administration

The CAA for Science field test content was delivered via the California Assessment of Student Performance and Progress (CAASPP) test delivery system (TDS). Authorized school and local educational agency (LEA) staff downloaded the *Directions for Administration (DFAs)* for each embedded PT from the secure Test Operations Management System (TOMS). Test examiners used the *DFA* materials in printed or electronic format.

#### Test Security and Confidentiality

All tests within the CAASPP System are secure. For the CAAs, every person having access to test materials maintains the security and confidentiality of the tests. ETS’ internal Code of Ethics requires that all test information, including tangible materials associated with the CAAs, confidential files, processes, and activities are kept secure. To ensure security for all tests that ETS develops or handles, ETS maintains an Office of Testing Integrity (OTI). A detailed description of the OTI and its mission is presented in subsection [*5.7.1 ETS’ Office of Testing Integrity*](#_ETS’_Office_of).

In pursuit of enforcing secure practices, ETS strives to safeguard the various processes involved in a test development and administration cycle. Those processes are listed in the following subsections and discussed in detail in [chapter 5](#_Test_Administration):

* [Standardization of test security](#_Procedures_to_Maintain)
* [Security of electronic files using a firewall](#_Security_of_Electronic)
* [Transfer of scores via secure data exchange](#_Transfer_of_Scores)
* [Data management](#_Data_Management_in)
* [Statistical analysis](#_Statistical_Analysis_on)
* [Student confidentiality](#_Student_Confidentiality)
* [Student test results](#_Test_Security_and)

#### Procedures to Maintain Standardization

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

##### Test Administration

ETS employs processes to ensure the standardization of an administration cycle; these processes are discussed in more detail in [*Chapter 5: Test Administration*](#_Test_Administration).

All staff at LEAs that are involved in the CAASPP administration, including CAA for Science administration, are provided directions about their responsibilities. Their roles include LEA CAASPP coordinators, CAASPP test site coordinators, and CAA test examiners. The responsibilities of each of the staff members specifically for the CAAs are described in the *2018–19 CAASPP Online Test Administration Manual* (CDE, 2019a).

##### Test Directions

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

* ***CAA for Science DFA*s—**A manual that provides the script and *DFAs* to be followed exactly by test examiners during a testing session. The secure *DFAs* for the CAA for Science contain item-specific instructions and therefore are grade- and version-specific. (Refer to [*5.4.4.1 Directions* *for* *Administration*](#_Directions_for_Administration) in [chapter 5](#_Test_Administration) for more information.)
* ***CAASPP Online Test Administration Manual*—**A manual that provides test administration procedures and guidelines for LEA CAASPP coordinators, CAASPP test site coordinators, test examiners, and test administrators (CDE, 2019a). (Refer to [*5.4.4.2 CAASPP* *Online* *Test* *Administration* *Manual*](#_CAASPP_Online_Test) in [chapter 5](#_Test_Administration) for more information.)
* ***TOMS Pre-Administration Guide for CAASPP Testing*—**A manual that provides instructions for TOMS that allow LEA staff, including LEA CAASPP coordinators and CAASPP test site coordinators, to perform a number of tasks including setting up test administrations, adding and managing users, and configuring online student test settings. (CDE, 2018b) (Refer to [*5.4.4.3 TOMS Pre-Administration Guide for CAASPP Testing*](#_TOMS_Pre-Administration_Guide) in [chapter 5](#_Test_Administration) for more information.)

### Test-Taking Rates

The decision to assign a student to take the CAA for Science is made by the student’s individualized education program (IEP) team, which uses the information on the Alternate Assessment IEP Team Guidance web page to make that determination. This web page describes the CAA and its administration, criteria for test takers, and the students who should be assigned to take this test (CDE, 2019b).

A student must meet all three of the following criteria to take the CAA for Science:

1. **The student has a significant cognitive disability.** Review of the student’s school records indicates a disability or multiple disabilities that significantly impact intellectual functioning and adaptive behavior essential for a person to live independently and to function safely in daily life.
2. **The student is learning content derived from the CA CCSS or the CA NGSS or is acquiring proficiency as identified in the 2012 English Language Development (ELD) Standards.** Goals and instruction listed in the IEP for the student are linked to the grade-level CA CCSS, CA NGSS, or 2012 ELD Standards and address knowledge and skills that are appropriate and set high expectations for this student.
3. **The student needs extensive, direct individualized instruction and substantial supports to achieve measurable gains in the grade-level and age-appropriate curriculum, including the following:**
   1. Instruction and support that are not of a temporary or transient nature
   2. Substantially adapted materials and individualized methods of accessing information in alternative ways to acquire, maintain, generalize, demonstrate, and transfer skills across multiple settings

All students who were identified to take the CAAs were required to test. All students were administered the Student Response Check (SRC) for each of the three embedded PTs. Students who did not provide a consistent, observable response to the SRC were not required to be administered the rest of the embedded PT.

Refer to [appendix 2.A](#_Appendix_2.A:_Test-Taking) for a summary of the number of test takers and the percent of the number of test takers for selected student groups for each test during the 2018–‍2019 administration. Because the data in the *Number of Test Takers* rows includes students with a valid test score, the number of test takers does not include students for whom the test was ended as a result of the SRC.

### Fairness and Accessibility

There are several procedures in place to ensure that the CAA for Science is fair and accessible to all test takers. This subsection provides information on the available accessibility resources for use with the online CAA for Science. Additionally, the differential item functioning (DIF) analysis used to identify items that may function differently across groups of examinees (e.g., gender, ethnicity) is also discussed briefly.

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

The CAAs are specifically designed for students with significant cognitive disabilities and an IEP that calls for the use of a CAA. Additional resources are sometimes needed for these students.

The CDE maintains a list of the universal tools, designated supports, and accommodations that are permitted for use in CAASPP online assessments in its web document “Matrix One: Universal Tools, Designated Supports, and Accommodations for the CAASPP System” (CDE, 2019c). [[4]](#footnote-5)

**Universal tools** are available to all students taking the CAA for Science. These resources may be turned on and off when embedded as part of the technology platform for the online CAA assessments on the basis of student preference and selection.

**Designated supports** are available to students taking the CAA for Science when determined as needed by an educator or team of educators, with parent/guardian and student input as appropriate, or when specified in the student’s IEP.

**Accommodations** must be permitted on the CAA for Science for all eligible students when specified in the student’s IEP.

While most of the resources presented for the CAASPP online assessments are accessible for the CAAs, there are a few resources that are not applicable because the CAAs are designed to be given one-on-one in the student’s language of instruction, using the student’s identified instructional resources.

For CAAs, designated supports and accommodations are assigned to individual students based on the needs identified through the student’s IEP. Such assignments are implemented in TOMS by the LEA CAASPP coordinator or CAASPP test site coordinator, either through individual assignment in the student’s profile in TOMS or by batch upload, where settings were uploaded into TOMS for multiple students. Settings were either selected and entered into a macro-enabled template—called the Individual Student Assessment Accessibility Profile (ISAAP) Tool—that created an upload file, or entered into a template. These designated supports and accommodations were delivered to the student through the TDS at the time of testing.

##### Selection of Accessibility Resources

The full list of the universal tools, designated supports, and accommodations that are used in CAASPP online assessments are documented in Matrix One (CDE, 2019c). Most embedded universal tools, designated supports, and accommodations listed in parts 1 and 2 of Matrix One are available for the CAA for Science through the online testing interface. Part 3 of Matrix One includes instructional and physical supports that are available for CAA for Science testing. School-level personnel and IEP teams use Matrix One when deciding how best to support the student’s test-taking experience. On the rare occasion when a student has both an IEP and a Section 504 plan, the Section 504 plan also should be referenced for accessibility resources.

In addition to assigning accessibility resources individually and via file upload in TOMS, LEAs had the option of using the ISAAP Tool to assign resources to students. The ISAAP Tool is used by LEAs in conjunction with the Smarter Balanced Assessment Consortium’s *Usability, Accessibility, and Accommodations Guidelines* (Smarter Balanced, 2019) and the Accessibility Guide for CAASPP Online Testing (CDE, 2019d), as well as with state regulations and policies (such as Matrix One) related to assessment accessibility*.*

##### Delivery of Accessibility Resources

Universal tools, designated supports, and accommodations can be delivered as either embedded or non-embedded resources. Embedded resources are digitally delivered features or settings available as part of the technology platform for the online CAAs. Examples of embedded resources applicable to the CAAs include masking, color contrast, and print size. Non-embedded resources for the CAAs include magnification, calculator, and scribe.

##### Unlisted Resources

An unlisted resource is an instructional support, identified in the student’s IEP, that a student regularly uses in daily instruction, assessment, or both, and has not been previously identified as a universal tool, designated support, or accommodation. Matrix One includes an inventory of unlisted resources that have already been identified and preapproved (CDE, 2019c). During the 2018–2019 CAASPP administration, an LEA CAASPP coordinator or a CAASPP test site coordinator had the option to submit a web form available in TOMS to request such a resource for an eligible student. The resource was required to be specified in the eligible student’s IEP and only assigned with the CDE’s approval.

For an unlisted resource to be approved, it must not change the construct of what is being tested for accountability purposes.

#### Individualizations

The CAA for Science is designed to strike a careful balance between standardized administration and maximizing student engagement. To meet this goal, some parts of each embedded PT can be individualized to improve student engagement. The individualizations are described in subsection [*5.5 Accessibility Features for the Field Test*](#_Accessibility_Features_for).

### Scores

Student responses to each embedded PT were scored by ETS through the TDS.

#### Score Reporting

There were no individual student scores reported for the 2018–2019 CAA for Science field test. ETS prepared an aggregate data file of students’ percent-correct scores and the associated preliminary indicator category for LEAs.

#### Aggregation Procedures

To provide meaningful results to the stakeholders, CAA for Science results for a given grade-level assessment were aggregated and generated 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 are presented for all students or selected demographic student groups.

The aggregation procedures used to present CAA for Science results are described in subsection [*6.2.3 Aggregate Score Reporting*](#_Aggregate_Score_Reporting). Aggregated scores that summarize student performance by grade for selected groups of students are provided in table 6.A.1 through table 6.A.3 in [appendix 6.A](#_Appendix_6.A:_Demographic). The tables show the numbers of students with valid scores in each group, raw score means and standard deviations (SDs), percent correct means and SDs, and percentage in an achievement level. Students are grouped by demographic characteristics, including gender, ethnicity, English-language fluency, primary disability, and economic status. Definitions for the demographic student groups included in these tables are provided in table 6.6.

### Overview of Psychometric Analyses

There were a number of psychometric analyses conducted on the CAA for Science results data. These are described fully in [*Chapter 7: Psychometric Analyses*](#_Psychometric_Analyses).

#### Description of the Classical Item Analyses

The psychometric analyses for the CAA for Science field test consisted of classical item analyses and DIF analyses to evaluate the performance of the embedded PT items. The classical item analyses include the computation of item difficulty indices, the item-total correlation indices, the omit rate of each embedded PT item, and the proportion of test takers obtaining each score point for the polytomous items. Flagging rules based on these statistics identify items not performing as expected. Descriptions of the psychometric analyses are provided in section [*7.2 Classical Item Analysis Statistics*](#_Toc22493011); [appendix 7.A](#_Appendix_7.A:_Classical) contains the results of the classical analyses*.*

Additionally, responses to the embedded PT survey questions were analyzed to evaluate how material choices and individualization might have impacted student performance. One unique aspect of the CAA for Science field test design was the flexibility offered to test examiners to exercise choice in the type of materials used. Test examiners could create testing conditions that were representative of classroom instruction, following the guidelines provided in the *DFAs*. Refer to [*Chapter 9: Embedded Performance Task and Test Comparability Considerations*](#_Embedded_Performance_Task)for details of these additional analyses*.*

#### Description of DIF Analyses

DIF analyses were conducted to detect differences in student performance by identifying items on which one group of students performs significantly better than another group (e.g., male vs. female or white vs. African-American) after matching students on ability. If an item performed differentially across student groups, even when students were matched on ability, the item may be measuring something other than the intended construct. Therefore, it is important to identify items flagged for DIF. Content experts and bias and sensitivity experts review these DIF-flagged items to determine the sources and meanings of performance differences. Refer to subsection [*7.5. DIF Analyses*](#_DIF_Analyses) for DIF analyses conducted and [appendix 7.D](#_Appendix_7.D:_DIF) for DIF analysis results.

#### Description of Item Response Theory (IRT) Analyses

A concurrent calibration was implemented to estimate parameters for all 2018–2019 field test items. As a result of the concurrent calibration, the item parameter estimates were placed on a common scale for test items from the same grade-level test. The concurrent calibration required either “common items” or “random equivalent groups.” The CAAs for Science versions were assembled with common items between the two versions, which supported the efficiency and accuracy of the concurrent calibrations. The one-parameter logistic IRT model (Hambleton and Rogers, 1991) and the partial credit model (Masters, 1982) were used for item calibration of the CAAs with flexMIRT® (Cai, 2016) version 3.5 software.

Detailed procedures for the concurrent calibrations are included in subsection [*7.6.2. Item Calibration*](#_Item_Calibration).

### References

Cai, L. (2016). FlexMIRT®: *Flexible multilevel, multidimensional item analysis and test scoring* (Version 3.5) [computer software]. Chapel Hill, NC: Vector Psychometric Group.

California Department of Education. (2018a). *California Alternate Assessment for Science blueprint*. Sacrament, CA: California Department of Education.

California Department of Education. (2018b). *TOMS pre-administration guide for CAASPP testing*. Sacramento, CA: California Department of Education.

California Department of Education. (2019d). *Accessibility guide for CAASPP online testing.* Sacramento, CA: California Department of Education.

California Department of Education. (2019b). *Alternate assessment IEP team guidance.*

California Department of Education. (2019a). *CAASPP online test administration manual, 2018–19 administration.* Sacramento, CA: California Department of Education.

California Department of Education. (2019c). *Matrix one:* *Universal tools, designated supports, and accommodations for the California Assessment of Student Performance and Progress for 2019–20.* Sacramento, CA: California Department of Education.

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

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–‍74.

Smarter Balanced Assessment Consortium. (2019). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations guidelines.* Los Angeles, CA: Smarter Balanced Assessment Consortium.

### Appendix 2.A: Test-Taking Rates

Table 2.A.1 CAA for Science Field Test Test-Taking Rates—Registered Students

|  |  |  |  |
| --- | --- | --- | --- |
| Test | Number of Registered Students | Number of Test Takers | Test Takers as a Percent of Registered Students |
| Grade 5 | 5,847 | 5,131 | 87.8 |
| Grade 8 | 6,000 | 5,217 | 87.0 |
| High school—Grade 10 | 521 | 377 | 76.6 |
| High school—Grade 11 | 3,471 | 2660 | 76.6 |
| High school—Grade 12 | 8,141 | 5,777 | 71.0 |
| High school—All grades | 12,133 | 8,814 | 72.6 |

Table 2.A.2 CAA for Science Field Test Test-Taking Rates for Test by Student Group, Grade Five

|  |  |  |  |
| --- | --- | --- | --- |
| Student Group | Number of Students | Number of Test Takers | Test Takers as a Percent of Registered Students |
| All students | 5,847 | 5,131 | 87.8 |
| Male | 3,935 | 3,432 | 87.2 |
| Female | 1,912 | 1,699 | 88.9 |
| American Indian or Alaska Native | 31 | 27 | 87.1 |
| Asian | 466 | 409 | 87.8 |
| Pacific Islander | 29 | 24 | 82.8 |
| Filipino | 137 | 122 | 89.1 |
| Hispanic or Latino | 3,282 | 2,962 | 90.2 |
| Black or African American | 479 | 419 | 87.5 |
| White | 1,148 | 943 | 82.1 |
| Two or more races | 224 | 185 | 82.6 |
| Ethnicity unknown | 51 | 40 | 78.4 |
| English only | 3,527 | 3,038 | 86.1 |
| Initial fluent English proficient | 59 | 57 | 96.6 |
| English learner | 1,793 | 1,600 | 89.2 |
| Reclassified fluent English proficient | 462 | 430 | 93.1 |
| To be determined | 0 | 0 | N/A |
| English proficiency unknown | 6 | 6 | 100.0 |
| Not economically disadvantaged | 2,011 | 1,677 | 83.4 |
| Economically disadvantaged | 3,836 | 3,454 | 90.0 |
| Intellectual disability | 2,038 | 1,813 | 89.0 |
| Hearing impairment | 53 | 48 | 90.6 |
| Speech or language impairment | 137 | 121 | 88.3 |
| Visual impairment | 24 | 18 | 75.0 |
| Emotional impairment | 33 | 23 | 69.7 |
| Orthopedic impairment | 238 | 196 | 82.4 |
| Other health impairment | 343 | 298 | 86.9 |
| Specific learning disability | 419 | 391 | 93.3 |
| Deaf-blindness | 4 | 3 | 75.0 |
| Multiple disabilities | 353 | 278 | 78.8 |
| Autism | 2,180 | 1,922 | 88.2 |
| Traumatic brain injury | 25 | 20 | 80.0 |
| Not classified | 0 | 0 | N/A |

Table 2.A.3 CAA for Science Field Test Test-Taking Rates for Test by Student Group, Grade Eight

|  |  |  |  |
| --- | --- | --- | --- |
| Student Group | Number of Students | Number of Test Takers | Test Takers as a Percent of Registered Students |
| All students | 6,000 | 5,217 | 87.0 |
| Male | 4,014 | 3,501 | 87.2 |
| Female | 1,986 | 1,716 | 86.4 |
| American Indian or Alaska Native | 35 | 29 | 82.9 |
| Asian | 444 | 398 | 89.6 |
| Pacific Islander | 32 | 29 | 90.6 |
| Filipino | 163 | 138 | 84.7 |
| Hispanic or Latino | 3,344 | 2,979 | 89.1 |
| Black or African American | 510 | 428 | 83.9 |
| White | 1,244 | 1,014 | 81.5 |
| Two or more races | 185 | 165 | 89.2 |
| Ethnicity unknown | 43 | 37 | 86.0 |
| English only | 3,521 | 2,983 | 84.7 |
| Initial fluent English proficient | 94 | 81 | 86.2 |
| English learner | 1,614 | 1,454 | 90.1 |
| Reclassified fluent English proficient | 767 | 696 | 90.7 |
| To be determined | 3 | 3 | 100.0 |
| English proficiency unknown | 1 | 0 | 0.0 |
| Not economically disadvantaged | 2,239 | 1,857 | 82.9 |
| Economically disadvantaged | 3,761 | 3,360 | 89.3 |
| Intellectual disability | 2,354 | 2,108 | 89.5 |
| Hearing impairment | 58 | 50 | 86.2 |
| Speech or language impairment | 101 | 83 | 82.2 |
| Visual impairment | 31 | 23 | 74.2 |
| Emotional impairment | 46 | 29 | 63.0 |
| Orthopedic impairment | 283 | 215 | 76.0 |
| Other health impairment | 329 | 279 | 84.8 |
| Specific learning disability | 417 | 375 | 89.9 |
| Deaf-blindness | 1 | 0 | 0.0 |
| Multiple disabilities | 338 | 257 | 76.0 |
| Autism | 2,008 | 1,771 | 88.2 |
| Traumatic brain injury | 34 | 27 | 79.4 |
| Not classified | 0 | 0 | N/A |

Table 2.A.4 CAA for Science Field Test Test-Taking Rates for Test by Student Group, High School

|  |  |  |  |
| --- | --- | --- | --- |
| Student Group | Number of Students | Number of Test Takers | Test Takers as a Percent of Registered Students |
| All students | 12,133 | 8,814 | 72.6 |
| Male | 7,987 | 5,832 | 73.0 |
| Female | 4,146 | 2,982 | 71.9 |
| American Indian or Alaska Native | 86 | 55 | 64.0 |
| Asian | 1,071 | 757 | 70.7 |
| Pacific Islander | 69 | 49 | 71.0 |
| Filipino | 436 | 308 | 70.6 |
| Hispanic or Latino | 6,191 | 4,602 | 74.3 |
| Black or African American | 1,021 | 710 | 69.5 |
| White | 2,843 | 2,038 | 71.7 |
| Two or more races | 335 | 240 | 71.6 |
| Ethnicity unknown | 81 | 55 | 67.9 |
| English only | 7,345 | 5,291 | 72.0 |
| Initial fluent English proficient | 236 | 154 | 65.3 |
| English learner | 2,830 | 2,049 | 72.4 |
| Reclassified fluent English proficient | 1,702 | 1,304 | 76.6 |
| To be determined | 8 | 6 | 75.0 |
| English proficiency unknown | 12 | 10 | 83.3 |
| Not economically disadvantaged | 5,108 | 3,550 | 69.5 |
| Economically disadvantaged | 7,025 | 5,264 | 74.9 |
| Intellectual disability | 5,143 | 3,893 | 75.7 |
| Hearing impairment | 107 | 74 | 69.2 |
| Speech or language impairment | 97 | 65 | 67.0 |
| Visual impairment | 87 | 57 | 65.5 |
| Emotional impairment | 109 | 55 | 50.5 |
| Orthopedic impairment | 690 | 441 | 63.9 |
| Other health impairment | 492 | 330 | 67.1 |
| Specific learning disability | 593 | 401 | 67.6 |
| Deaf-blindness | 4 | 3 | 75.0 |
| Multiple disabilities | 815 | 532 | 65.3 |
| Autism | 3,897 | 2,895 | 74.3 |
| Traumatic brain injury | 99 | 68 | 68.7 |
| Not classified | 0 | 0 | N/A |

## Embedded PT Development and Review

This chapter provides an overview of the processes implemented by Educational Testing Service (ETS) to develop items for use on the California Alternate Assessment (CAA) for Science. These processes include those that are entirely internal to ETS and those that are conducted in coordination with the California Department of Education (CDE).

The chapter provides a brief description of each process and a summary of the associated specifications. More details about the specifications and the analyses associated with each process are described in other chapters that are referenced in the subsections that follow.

### Embedded PT and Item Development

Each CAA for Science embedded performance task (PT) item is developed through a comprehensive cycle and designed to conform to ETS-defined principles of item writing. Each item in the CAA for Science item bank was developed to measure a specific California Next Generation Science Standard (CA NGSS) Core Content Connector (Science Connector). The Science Connectors are based on the performance expectations (PEs) from the CA NGSS and were designed to incorporate the science and engineering practices, disciplinary core ideas, and the crosscutting concepts that comprise the CA NGSS. The Science Connectors are further broken down into more discrete focal knowledge, skills, and abilities (FKSAs) and, at the simplest level, the essential understandings (EUs). In addition, guidelines for style, fairness, and bias and sensitivity help item developers and reviewers ensure consistency across the item development process.

#### Specifications for the Embedded PTs and Items

The item specifications for prioritized Science Connectors describe the characteristics of the tasks developed to measure each Science Connector and provide detailed information to task writers who develop items for the CAA for Science. The specifications include the following:

* The full statement of the associated CA NGSS PE
* The full statement of the Science Connector
* The full content of each assessed FKSA of the Science Connector
* The full content of each assessed EU of the Science Connector
* How mastery of the EUs and FKSA(s) is demonstrated

#### Embedded PT and Item Format

Embedded PTs for the CAA for Science were designed to be engaging to the target population. Embedded PTs were developed with the understanding that a test examiner would deliver each task individually to each eligible student and assist the student in responding as appropriate during each portion of the embedded PT. Instructions and guidance for each embedded PT are contained within the embedded PT *DFA*.

Each embedded PT *DFA* began with background information and instructions for the test examiner. These instructions included the following:

* student engagement, student response, and survey;
* the concept of individualization;
* Student Response Decision Matrix (refer to subsection [*5.3.1 Administration of the Student Response Check*](#_Administration_of_the))
* orienting activities and graphics for the orienting activities, if needed;
* the associated script for the online test questions, and
* a complete list of materials needed for the administration of this embedded PT and suggestions for individualization, if needed.

The CAA for Science included the following item formats:

* **Selected Response—**Students were instructed to select one or more choices. Most CAA for Science items had two or three options; a few items have four options.
* **Match—**Students were instructed to place a picture on a specified part of a diagram or chart.
* **Grid—**Students were instructed to place a checkmark in a specified section in a table of responses.

All selected-response and match items were scored by the test delivery system.

The number of items and points for each embedded PT is provided in table 3.1.

Table 3.1 Number of Items and Points for Each Embedded PT

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Module | Number of Items—PT 1 | Number of Points—PT 1 | Number of Items—PT 2 | Number of Points—PT 2 | Number of Items—PT 3 | Number of Points—PT 3 | Total Number of Items | Maximum Number of Points |
| Grade 5 | 10 | 12 | 10 | 12 | 10 | 12 | 30 | 36 |
| Grade 8 | 10 | 12 | 10 | 12 | 10 | 12 | 30 | 36 |
| High School | 10 | 12 | 10 | 12 | 10 | 12 | 30 | 36 |

#### Recruitment and Selection of Embedded PT Item Writers

Applications for embedded PT item writing were screened by senior ETS content staff. Only those applicants with strong science content or special education teaching backgrounds were approved for inclusion in the training program for item writing.

All item writers met the following minimum qualifications:

* Possession of a bachelor’s degree in a science content area or in the field of education with special focus on a particular science content area; an advanced degree in science or special education is desirable
* Experience teaching students with cognitive disabilities and preferably experience teaching science in grades five through twelve
* Previous experience or training in writing items for standards-based assessments, including knowledge of the many considerations that are important when developing items for special student populations
* Previous experience or training in writing items in the content areas covered by CAA grades, content domains, or both
* Familiarity, understanding, and support of the Science Connectors, EUs, and FKSAs

#### Embedded PT Item Writer Training

Item writer training for the field test cycle took place over two days in July 2018. Attendees received training on the Science Connectors used for the CAA for Science, general principles of universal design, CAA for Science item specifications, and how to account for bias and sensitivity when writing items.

During the training, attendees wrote sample items that were evaluated and returned with feedback from ETS science assessment specialists.

### Embedded PT and Item Review Process

#### Selection of Embedded PTs and Items

The activities and items developed for the CAA for Science embedded PTs underwent an extensive item review process that was designed to provide the best standards-based assessments possible. This subsection summarizes the item review process that ensured the quality of CAA for Science activities and items.

Tasks and items submitted by the item writers were reviewed by ETS assessment specialists, who determined whether or not each embedded PT and item met the criteria expected for submission, including accuracy and adherence to the item specifications. Embedded PTs and items that did not meet the criteria were rejected, with notes for future revision submitted to authors. Items that met the criteria were accepted into the pool and authored into the system.

Once an item was accepted for further development—that is, once it was entered into the ETS item bank and formatted for use in an assessment—ETS employed a series of internal reviews to judge the quality of item content and ensure that each item measured what it was intended to measure. These internal reviews also examined the overall quality of the test items before presentation to the CDE and California educators.

The ETS review process for the CAA for Science included the following; these are described in the next subsections:

1. Content review
2. Editorial review
3. Sensitivity review

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

#### ETS Content Review

Embedded PTs underwent three rounds of content reviews by content-area assessment specialists with increasing levels of expertise, called Round 1, Round 2, and Final Round. The assessment specialists ensured thatthe embedded PTs complied with the approved item specifications and with ETS written guidelines for clarity, style, accuracy, and appropriateness for California students. Assessment specialists reviewed each embedded PT and item for the following characteristics:

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

Each embedded PT item was classified with the Science Connector, EU, and the FKSA it was intended to measure. Assessment specialists checked each item against its classification codes, both to evaluate the correctness of the classification and to ensure 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 Editorial Review

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

#### ETS Sensitivity and Fairness Review

ETS assessment specialists who are specially trained to identify and eliminate questions that contain content or wording that could be construed to be offensive to or biased against members of specific student groups—ethnic, racial, or gender—conducted the next level of review. These trained staff members reviewed every item before the CDE and formal embedded PT item reviews.

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

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

### Content Expert Review

#### California Educators as Content Experts

Meetings with California educators were held at the end of the item review process as the final content expert review that items must undergo before being placed on an operational assessment. The California educators filled an advisory role to the CDE and ETS and provided guidance on matters related to embedded PT item development for the CAA for Science.

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

#### Composition of Item Review Panels

The group of California educators participating in item review meetings consisted of current and former teachers (some of whom taught students who comprised the identified population, and others who were subject matter experts in science), resource specialists, administrators, curriculum and content experts, and other education professionals. Minimum qualifications to be invited to participate were

* three or more years of teaching experience in kindergarten through grade twelve, and
* bachelor’s or higher degree in a grade or content area related to special education.

Preferred qualifications included

* experience teaching students with more than one type of disability, and
* three to five years of experience as a teacher or school administrator with a special education credential.

School administrators; local educational agency (LEA), county content, or program specialists; or university educators must have met the following qualifications to be invited to participate:

* Three or more years of experience as a school administrator; LEA, county content, or program specialist; or university instructor in a content-specific area; and
* Knowledge of and experience with the CA NGSS.

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

Table 3.2 shows the educational qualifications, present occupation, and credentials of the individuals who participated in CAA for Science item review. Note that some reviewers had multiple occupations or teaching credentials, and some were working toward earning their highest degree.

Table 3.2 Number of Item Reviewers with Each Qualification

| **Qualification Type** | **Qualification** | **Number of Reviewers** |
| --- | --- | --- |
| **N/A** | **Total number of reviewers** | **10** |
| **Occupation** | Teacher or Program Specialist, Elementary School | 3 |
| **Occupation** | Teacher or Program Specialist, Middle School | 4 |
| **Occupation** | Teacher or Program Specialist, High School | 3 |
| **Occupation** | Other District Personnel | 0 |
| **Highest Degree Earned** | Bachelor’s Degree | 0 |
| **Highest Degree Earned** | Master’s Degree | 8 |
| **Highest Degree Earned** | Doctorate | 9 |
| **K–12 Teaching Credential** | Elementary Teaching (multiple subjects) | 4 |
| **K–12 Teaching Credential** | Secondary Teaching (single subject) | 2 |
| **K–12 Teaching Credential** | Special Education | 4 |
| **K–12 Teaching Credential** | Reading Specialist | 1 |
| **K–12 Teaching Credential** | English Learner (CLAD, BCLAD) | 0 |
| **K–12 Teaching Credential** | Administrative | 0 |
| **K–12 Teaching Credential** | Other | 0 |

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

#### Meetings for Review of CAA for Science Embedded PTs and Items

ETS science assessment specialists facilitated CAA for Science item review meetings. Each meeting began with a brief training session on how to review embedded PT items. ETS provided this training, which consisted of the following topics:

* Overview of the purpose and scope of the CAA for Science
* Overview of the CAA for Science test design specifications and blueprints
* Analysis of the CAA for Science embedded PT item specifications
* Overview of criteria for evaluating test items
* Review and evaluation of items for bias and sensitivity issues

The criteria for evaluating items included the following:

* Overall technical quality
* Match to the Science Connectors
* Match to the construct being assessed by the Science Connector
* Difficulty range
* Clarity
* Correctness of the answer
* Plausibility of the distractors
* Bias and sensitivity factors

Criteria also encompassed more global factors, including the quality of the alternative text to confirm that it describes an image in an age- and audience-appropriate manner within the context of the question. Meeting participants also were trained on how to make recommendations for revising items.

Guidelines for reviewing items were provided by ETS and approved by the CDE. The set of guidelines for reviewing items is summarized next.

* Does the item
* have one and only one clearly correct answer (for single-select items)?
* measure the content standard?
* match the item specifications?
* align with the construct being measured?
* test worthwhile concepts or information?
* Is the stimulus, if any, for the item
* required in order to answer the item?
* likely to be interesting to students?
* clearly and correctly labeled?
* providing all the information needed to answer the item?

### Data Review Meeting

After items were administered to students, ETS prepared the items and the associated statistics for review by the CDE and California educators.

For the CAA for Science, review materials included embedded PT items with their statistical data along with annotated comment sheets for use by reviewers. ETS conducted an introductory training to highlight any new issues and serve as a statistical refresher. Reviewers then made decisions about which items should be included in the item bank for future assembly. If an item was considered problematic and not to be included in the item bank, it would be revised and once again followed 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 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 the item in the operational pool.

## Test Assembly

This chapter provides details of test assembly, including a description of the content being measured (i.e., test blueprints), process of item selection, final reviews before test production, and the production process (e.g., preparation of the test forms for online test delivery).

### Test Content Specifications and Test Blueprints

The California Alternate Assessment (CAA) for Science incorporates innovations and best practices from recent national alternate assessment initiatives, including the National Center and State Collaborative and the Dynamic Learning Maps. All items and tasks are developed to the California Next Generation Science Standards Core Content Connectors (Science Connectors) developed by California educators, Educational Testing Service (ETS), and EdCount. An essential understanding (EU) and focal knowledge, skills, and abilities (FKSA) are identified for each Science Connector. EUs define a basic, foundational key idea or concept based on the Science Connector that builds increasing understanding of the grade-level content. FKSAs provide more specific detail about the requirements described by the Science Connectors.

#### Test Content Specifications

The CAA for Science assesses each Science Connector through the FKSAs and EUs derived from the Science Connectors. These Science Connectors identify the most salient grade-level, core academic content in science found in the CA NGSS, and illustrate the necessary knowledge and skills required to reach the learning targets within the CA NGSS. Additionally, the Science Connectors focus on the core content, knowledge, and skills needed to help students at each grade level succeed; and identify priorities in science to guide the instruction for students in this population and for an alternate assessment. Finally, the Science Connectors provide a foundation that permits teachers, parents/guardians, and the students themselves to help students with significant cognitive disabilities identify and address gaps in knowledge or skills early so students can receive the support they need (California Department of Education [CDE], 2020).

Each content standard is assessed through the Science Connectors and related FKSAs and EUs under a three-level structure of item complexity. Detailed information on the levels of cognitive complexity is provided in subsection [*4.2 Test Design*](#_Test_Design).

#### Test Blueprints

The CAA for Science test blueprints are unique to each grade band or level (California Department of Education [CDE], 2018). These blueprints designate the breakdown of each assessment, first by science domain and then by Science Connectors. Information on a test blueprint for a given grade and content area includes the

* specific ratio of each content domain on the overall test;
* specific Science Connectors to be assessed; and
* number of items on a test.

The 2018–2019 forms had 100 percent alignment with the test blueprint. Each of the three content domains were assessed by 10 items for a total of 12 points.

Overall, the percent of items per content domain based on the Science Connector assigned during item development and those in the CAA blueprint are comparable.

### Test Design

#### Operational Field Test

The field test was administered as an operational field test with the goal of building a bank of operational items. Three embedded performance tasks were administered during the year in each of grades five and eight and in high school. Each embedded performance task within a grade assessed one of the three science domains. These domains are Life Sciences, Physical Sciences, and Earth and Space Sciences.

Each embedded performance task assessed two Science Connectors from a domain. The embedded performance tasks contained an orienting activity and five questions aligned to each of the two Science Connectors. Thus, an embedded performance task contained two orienting activities and 10 questions. Some of the orienting activities and questions incorporated simple activities designed to demonstrate a key concept associated with the assessed Science Connector.

### Test Production Process

#### Selection of Items

From the eligible item pool, test developers selected items that, as a whole,

* met the coverage specifications of the test blueprint,
* met the form-building guidelines developed by the ETS psychometrics team,
* represented a wide variety of item types, and
* provided a wide variety of item context.

#### Content Review of Forms

After psychometric approval, the proposed assessment underwent two additional content reviews and one editorial review. The form reviewers are content specialists who work on testing programs other than the CAA for ETS, so they were able to bring a fresh perspective to the review. They are given the appropriate materials to do the following:

* Verification of item keys
* Identification of possible clueing across the items
* Verification that individual items met the standard
* Verification of coverage of the standards
* Identification of any possible grammatical or production errors

#### CDE Forms Review

Following the ETS content review, all proposed assessments were sent to the CDE for review to ensure the proposed assessments met CAA for Science test blueprint requirements and to check there was no clueing between items. The CDE was provided with the following materials:

* Access to items in the item banking system
* Modified form planners
* Comment sheets

Comments from the CDE were resolved during a virtual meeting with the ETS test development team.

#### Configuration of the TDS

Once all the test reviews were completed and concerns, if any, were resolved, the official ordered item sequence of the proposed forms was sent to the American Institutes for Research (AIR) for configuration of the California Assessment of Student Performance and Progress test delivery system (TDS).

AIR’s TDS supported a variety of item layouts. Some of the item layouts had the stimulus and item response options and response area displayed side by side. In each of these item layouts, both the stimulus and response options had independent scroll bars. 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 AIR consisting of a team leader and several team members. The team leader presented the item as it was approved in ETS and AIR item banks. Each team member was assigned a different platform—hardware device and operating system—and reviewed the item to verify that it rendered as expected. This platform review meeting ensured that all items would be 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 subjected to user acceptance testing (UAT) by both ETS and AIR staff. The TDS UAT served as both a software evaluation and a content approval role.

The UAT procedures followed by the ETS staff included reviewing all items.

Following the UAT by ETS and AIR 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 CAA UAT before the test could be released for administration to students.

### References

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacrament, CA: California Department of Education.

California Department of Education. (2020). *California Alternate Assessment for Science*. Sacramento, CA: California Department of Education.

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

## Test Administration

This chapter describes the administration of the embedded performance tasks (PTs) for the California Alternate Assessment (CAA) for Science field test, as well as the procedures followed by Educational Testing Service (ETS) to ensure test security.

### Grade Assignment

All local educational agencies (LEAs) with eligible students in grades five and eight and high school (grade ten, eleven, or twelve) administered the field test for the CAA for Science. Students in high school who were selected by the LEA to take a science assessment and whose individualized education program (IEP) indicated an alternate assessment were assigned to take the CAA for Science (California Science Teachers Association, 2000–2019).

Students in grades five and eight and in high school (grade ten, eleven, or twelve) who met all of the following eligibility requirements took the CAA for Science:

* The student has a significant cognitive disability that is described in the student’s IEP.
* The student is learning content derived from the California Next Generation Science Standards Core Content Connectors (Science Connectors).
* The student requires extensive direct individualized instruction and substantial resources to achieve measurable gains in the grade- and age-appropriate curriculum.

### Administration Preparations

The embedded PTs were designed to be administered to students in conjunction with the normal course of instruction related to the Science Connector being assessed. The test examiner was instructed to administer the embedded PT shortly after the student received instruction aligned with the Science Connector.

#### Guides and Videos

To supplement the in-person workshops and the live webcast, ETS also produced short “how-to” videos and narrated PowerPoint presentations that were available on the California Assessment of Student Performance and Progress (CAASPP) Summative Assessments Training Videos web page. For the CAA for Science, two videos were produced.

The first video, “Administering the California Alternate Assessment for Science,” included the following topics and demonstrations (California Department of Education [CDE], 2020):

* Overview of 2018–2019 CAA for Science field test
* Checklist of activities prior to administering a test
* How to download the embedded PT
* How the embedded PT is organized
* How to individualize, task scoring
* Other available resources

ETS also produced an online module, the CAA Test Examiner Tutorial, designed to teach test examiners how to administer the CAAs, including the CAA for Science (CDE, 2019a). Test examiners were required to complete a training session before administering the CAAs by either completing a local training or completing this stand-alone online training module.

#### Local Educational Agency (LEA) Training

ETS established and implemented a training plan for LEA assessment staff on all aspects of the assessment program. The CDE and ETS, in collaboration with stakeholders as needed, determined the audience, topics, frequency, and mode (in-person, webcast, videos, modules, etc.) of the training, including such elements as format, participants, and logistics.

ETS conducted 16 in-person pretest workshops and presented four webcasts for the 2018–‍2019 administration. Additionally, ETS produced a tutorial for CAA administration.

Following approval by the CDE, the ancillary materials were posted for each webcast on the CAASPP website so the LEAs could download the training materials.

##### In-person Training

ETS provided a series of in-person trainings. Beginning in January 2019, the first in-person trainings provided were the pretest workshops, which focused on training LEA CAASPP coordinators on how to prepare for administering the CAASPP. Additionally, a two-session Post-Test Workshop was offered in May and June 2019 with the sessions “Principles of Scoring and Reporting” and “The Results Are In—Now What?”

##### Webcasts

ETS provided a series of live webcasts throughout the school year that were archived and made available for training LEA and test site staff as well as test examiners. Webcast viewers were provided with a method of electronically submitting questions to the presenters during the webcast. The webcasts were recorded and archived for on-demand viewing on the CAASPP Summative Assessments Training Videos web page. CAASPP webcasts were available to everyone and required neither preregistration nor a logon account.

##### Videos and Narrated PowerPoint Presentations

To supplement the in-person workshops and the live webcast, ETS also produced short “how-to” videos and narrated PowerPoint presentations that were available on the CAASPP Summative Assessments Training Videos web page.

Finally, ETS produced an online module, the CAA Test Examiner Tutorial, designed to teach test examiners how to administer the CAA for Science. Test examiners are required to complete a training session before administering the CAAs by either completing a local training or completing this stand-alone online training module (CDE, 2019a).

### Test Administration

The CAA for Science field test was administered one-on-one by a test examiner familiar with the student being tested. The test examiner administered three embedded PTs to each student; these were administered online through the California Assessment of Student Performance and Progress (CAASPP) test delivery system (TDS).

#### Administration of the Student Response Check

Prior to beginning the embedded PT, the test examiner conducted a Student Response Check (SRC) with the student to verify whether the student had a consistent and observable way of indicating responses to test questions. Student response modes may include indicating an answer with a mouse or keyboard, verbalizations, pointing, or gesturing. Students also may respond using eye gaze and an assistive communication device.

For the field test, test examiners conducted an SRC with the student at the start of each embedded PT administration. Each embedded PT provided instructions to the test examiner to use objects from the materials list for the particular embedded PT. The test examiner showed the objects to the student and directed the student to identify one familiar object in the set of objects, using the student’s mode of communication. For example, the test examiner might say, “Show me the flashlight.” If the student communicated an observable response, even if the selection was incorrect, the text examiner administered the embedded PT. If the student did not communicate an observable response, the test examiner did not administer the embedded PT.

#### Administration of the Embedded PTs

The embedded PTs were designed to be administered to students in conjunction with the normal course of instruction related to the Science Connector being assessed. The test examiner was instructed to administer the embedded PT shortly after the student received instruction related to the Science Connector.

#### Administration of the Survey

After an embedded PT was administered to a student, the test examiner answered four questions about any individualizations used during the test administration and one question about the level of student engagement. The results of these survey questions were entered into the TDS.

Additionally, an optional test examiner survey available on caaspp.org was used to solicit feedback regarding the test examiner’s experience with the assessment. Refer to [*8.2 Test Examiner Survey Results*](#_Test_Examiner_Survey) for more information about the feedback received.

### Procedures to Maintain Standardization

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

#### LEA California Assessment of Student Performance and Progress (CAASPP) Coordinator

An LEA CAASPP coordinator was designated by the district superintendent at the beginning of the 2018–2019 school year. LEAs include public school districts, statewide benefit charter schools, State Board of Education–authorized charter schools, county office of education programs, and direct funded charter schools.

LEA CAASPP coordinators are responsible for ensuring the proper and consistent administration of the assessments that are part of the CAASPP System, including the CAAs. In addition to the responsibilities set forth in the *California Code of Regulations*, Title 5 (5 *CCR*) Section 857, their responsibilities include

* adding CAASPP test site coordinators and test examiners into the Test Operations Management System (TOMS);
* training CAASPP test site coordinators and test examiners regarding state requirements and CAA administration as well as security policies and procedures;
* reporting test security incidents (including testing irregularities) to the CDE;
* overseeing test administration activities;
* filing a report of a testing incident in the Security and Test Administration Incident Reporting System (STAIRS); and
* requesting an Appeal (if indicated by TOMS prompts while reporting an incident using the STAIRS/Appeal process).

#### CAASPP Test Site Coordinator

A CAASPP test site coordinator is trained by the LEA CAASPP coordinator or district superintendent 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 CAASPP test site coordinator is responsible for identifying test examiners and ensuring that they have signed CAASPP Test Security Affidavits (5 *CCR* Section 859[d]). CAASPP test site coordinators’ duties may include

* adding test examiners into TOMS;
* entering test settings for students;
* creating testing schedules and procedures for a school consistent with state and LEA policies;
* working with technology staff to ensure secure browsers are installed and any technical issues are resolved;
* monitoring testing progress during the testing window and ensuring all students take the test, as appropriate;
* coordinating and verifying the correction of student data errors in the California Longitudinal Pupil Achievement Data System;
* ensuring a student’s test session is rescheduled, if necessary;
* addressing testing problems;
* reporting security incidents;
* overseeing administration activities at a school site;
* 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/Appeal process).

#### Test Examiners

Test examiners are identified by CAASPP test site coordinators as individuals who will administer the CAASPP assessments, including the CAA for Science. A test examiner must be a certificated or licensed school staff member (5*CCR* Section 850[ag]) and sign a security affidavit (5 *CCR* Section 859[d]).

A test examiner’s duties may include

* participating in training by either viewing the online test administration tutorial or attending any locally provided training;
* ensuring the physical conditions of the testing room meet the criteria for a secure test environment;
* administering the CAAs;
* reporting all test security incidents to the test site coordinator and LEA CAASPP coordinator in a manner consistent with state and LEA policies;
* viewing student information prior to testing to ensure that the correct student receives the proper test with appropriate resources and reporting potential data errors to test site coordinators and LEA CAASPP coordinators;
* monitoring student progress throughout the test session using the Test Administrator Interface; and
* complying fully with all directions provided in the Directions for Administration (DFA) for the CAA for Science.

#### Instructions for Test Examiners and Staff Involved in CAA for Science Administration

##### *Directions for Administration*

Test examiners used the *Embedded Performance Task Directions for Administration for the CAA for Science* to administer each separate embedded PT to students. The *DFAs* included the description of the activity, list of the exemplar materials, and the exemplar script. *DFAs* also included scoring rubrics where warranted.

Sample *Directions for* *Administration for the California Alternate Assessments* to be used in conjunction with the CAA practice and training tests were provided to LEAs as well (CDE, 2019b).

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

The *CAASPP Online Test Administration Manual* (CDE, 2019c) contains information and instructions on overall procedures and guidelines for all LEA and test site staff involved in the administration of online assessments as well as for the CAA for Science. 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
* Instructions for steps to take before, during, and after testing

Appendices included definitions of common terms, item types, descriptions of different aspects of the test and systems associated with the test, and checklists of activities for LEA CAASPP coordinators, CAASPP test site coordinators, and test examiners.

##### *TOMS Pre-Administration Guide for CAASPP Testing*

TOMS is a web-based application that allows LEA CAASPP coordinators to set up test administrations, add and manage users, and submit online student test settings. Test examiners accessed TOMS to retrieve *CAA for Science DFAs*.

TOMS modules included the following (CDE, 2018a):

* **Test Administration Setup—**This module allowed LEAs to determine and calculate dates for the LEA’s 2018–2019 testing.
* **Adding and Managing Users—**This module allowed LEA CAASPP coordinators to add CAASPP test site coordinators and test examiners to TOMS so that the designated user could access the online embedded PT *DFAs*.
* **Student Test Assignment—**This module allowed LEA CAASPP coordinators to designate students to take the alternate assessments.

##### *CAA for Science Administration Planning Guides*

The *Administration Planning Guides*, posted prior to the annual launch of the embedded PTs, provided information about the embedded PTs that will be administered in the coming school year (CDE, 2019d). The *Administration Planning Guides* contained information to help test examiners understand how to plan for the administration of the embedded PTs throughout the school year, version assignments, and test security. The *Administration Planning Guides* also contained the following information:

* Questions and answers about administration
* Task standards table
* How mastery of the Science Connector is demonstrated

##### 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.

* ***Technical Specifications and Configuration Guide for CAASPP Online Testing*—**This manual provides information, tools, and recommended configuration details to help technology staff prepare computers and install the secure browser to be used for the online CAASPP assessments (CDE, 2018b).
* ***Security Incidents and Appeals Procedure Guide*—**This manual provides information on how to report a testing incident and submit an Appeal to the CDE to reset, reopen, invalidate, or restore individual online student assessments (CDE, 2019e).

### Accessibility Features for the Field Test

#### Individualizations

A notable feature of the 2018–2019 embedded PTs is that test examiners had the option to individualize certain elements of the assessment, although not all embedded PTs allowed for individualization. For the field test administration, test examiners were instructed to review the activities associated with each embedded PT and decide whether the exemplar activity met a student’s needs or if an individualized activity was appropriate. The test examiner documented the use of individualizations in the survey at the end of each embedded PT.

Potential individualizations were designed so that the premise of the item and the scientific principles tested would remain the same. Individualization options in embedded PTs often involved the use of objects to make certain science concepts easier to understand for some students.

Table 5.1 through table 5.3 display the results of the survey regarding the kinds of individualization provided. N‑counts in these tables are based on all students in version 2 of the production file (“P2”) released on August 30, 2019, with an include indicator of “T” to indicate the student tested. Although test examiners are permitted to individualize the administration of the CAA for Science, table 5.1 through table 5.3 indicate that few students received individualizations, meaning the majority of students were administered the embedded PTs as outlined in the *DFA*s.

Table 5.1 Individualizations**—Grade Five**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Individualization | PT 1 (Life Sciences), Activity 1—Number | PT 1 (Life Sciences), Activity 1—Percent of Total | PT 1 (Life Sciences), Activity 2—Number | PT 1 (Life Sciences), Activity 2—Percent of Total | PT 2 (Physical Sciences), Activity 1—Number | PT 2 (Physical Sciences), Activity 1—Percent of Total | PT 2 (Physical Sciences), Activity 2—Number | PT 2 (Physical Sciences), Activity 2—Percent of Total | PT 3 (Earth and Space Sciences), Activity 1—Number | PT 3 (Earth and Space Sciences), Activity 1—Percent of Total | PT 3 (Earth and Space Sciences), Activity 2—Number | PT 3 (Earth and Space Sciences), Activity 2—Percent of Total |
| Using Standardized Scripts | 5,824 | 100% | 5,827 | 100% | 5,820 | 100% | 5,824 | 100% | 5,821 | 100% | 5,817 | 99% |
| Using Individualized Scripts | 23 | 0% | 20 | 0% | 27 | 0% | 23 | 0% | 26 | 0% | 30 | 1% |
| Using Standardized Graphic | 5,819 | 100% | 5,802 | 99% | 5,822 | 100% | 5,839 | 100% | 5,824 | 100% | 5,836 | 100% |
| Using Individualized Graphic | 28 | 0% | 45 | 1% | 25 | 0% | 8 | 0% | 23 | 0% | 11 | 0% |
| Using Standardized Materials | 5,777 | 99% | 5,819 | 100% | 5,814 | 99% | 5,702 | 98% | 5,834 | 100% | 5,770 | 99% |
| Using Individualized Materials | 70 | 1% | 28 | 0% | 33 | 1% | 145 | 2% | 13 | 0% | 77 | 1% |

Table 5.2  **Individualizations—Grade Eight**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Individualization | PT 1 (Life Sciences), Activity 1—Number | PT 1 (Life Sciences), Activity 1—Percent of Total | PT 1 (Life Sciences), Activity 2—Number | PT 1 (Life Sciences), Activity 2—Percent of Total | PT 2 (Physical Sciences), Activity 1—Number | PT 2 (Physical Sciences), Activity 1—Percent of Total | PT 2 (Physical Sciences), Activity 2—Number | PT 2 (Physical Sciences), Activity 2—Percent of Total | PT 3 (Earth and Space Sciences), Activity 1—Number | PT 3 (Earth and Space Sciences), Activity 1—Percent of Total | PT 3 (Earth and Space Sciences), Activity 2—Number | PT 3 (Earth and Space Sciences), Activity 2—Percent of Total |
| Using Standardized Scripts | 5,980 | 100% | 5,980 | 100% | 5,973 | 100% | 5,974 | 100% | 5,982 | 100% | 5,980 | 100% |
| Using Individualized Scripts | 20 | 0% | 20 | 0% | 27 | 0% | 26 | 0% | 18 | 0% | 20 | 0% |
| Using Standardized Graphic | 5,963 | 99% | 5,973 | 100% | 5,984 | 100% | 5,992 | 100% | 5,986 | 100% | 5,992 | 100% |
| Using Individualized Graphic | 37 | 1% | 27 | 0% | 16 | 0% | 8 | 0% | 14 | 0% | 8 | 0% |
| Using Standardized Materials | 5,971 | 100% | 5,979 | 100% | 5,845 | 97% | 5,854 | 98% | 5,911 | 99% | 5,988 | 100% |
| Using Individualized Materials | 29 | 0% | 21 | 0% | 155 | 3% | 146 | 2% | 89 | 1% | 12 | 0% |

Table 5.3  **Individualizations—High School**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Individualization | PT 1 (Life Sciences), Activity 1—Number | PT 1 (Life Sciences), Activity 1—Percent of Total | PT 1 (Life Sciences), Activity 2—Number | PT 1 (Life Sciences), Activity 2—Percent of Total | PT 2 (Physical Sciences), Activity 1—Number | PT 2 (Physical Sciences), Activity 1—Percent of Total | PT 2 (Physical Sciences), Activity 2—Number | PT 2 (Physical Sciences), Activity 2—Percent of Total | PT 3 (Earth and Space Sciences), Activity 1—Number | PT 3 (Earth and Space Sciences), Activity 1—Percent of Total | PT 3 (Earth and Space Sciences), Activity 2—Number | PT 3 (Earth and Space Sciences), Activity 2—Percent of Total |
| Using Standardized Scripts | 12,105 | 100% | 12,104 | 100% | 12,101 | 100% | 12,103 | 100% | 12,104 | 100% | 12,103 | 100% |
| Using Individualized Scripts | 28 | 0% | 29 | 0% | 32 | 0% | 30 | 0% | 29 | 0% | 30 | 0% |
| Using Standardized Graphic | 12,112 | 100% | 12,109 | 100% | 12,129 | 100% | 12,131 | 100% | 12,115 | 100% | 12,112 | 100% |
| Using Individualized Graphic | 21 | 0% | 24 | 0% | 4 | 0% | 2 | 0% | 18 | 0% | 21 | 0% |
| Using Standardized Materials | 11,993 | 99% | 12,097 | 100% | 11,733 | 97% | 11,965 | 99% | 11,992 | 99% | 12,068 | 99% |
| Using Individualized Materials | 140 | 1% | 36 | 0% | 400 | 3% | 168 | 1% | 141 | 1% | 65 | 1% |

#### Choice of Administration Scripts

Test examiners had the option of using an individualized script different than the suggested exemplar script to improve engagement for students who otherwise may not engage at all with the activity or item.

#### Choice of Materials

The activities that are part of each embedded PT were almost all designed to allow test examiners to substitute different materials as long as the required Science Connector activity was administered. Test examiners were permitted to substitute different materials based on the needs of the student as long as the purpose of the activity was followed. Suggested choices were designed so the scientific principles tested would remain the same. Embedded PTs often involved the use of objects to scaffold—build on the concepts to make them easier to understand—the scientific principles.

For example, for a particular high school embedded PT, before test questions regarding erosion were asked, test examiners were instructed to administer an exemplar activity that used soil, aquarium gravel, and water to demonstrate the effects of water on the Earth’s materials and surface processes. Test examiners had the option of substituting the exemplar materials with other materials listed (e.g., “small rocks, gravel, or metal BBs” in place of aquarium gravel and “sand or cornmeal” in place of soil).

#### Type and Level of Accommodations

For the administration of the embedded PTs, teachers were guided to offer the same instructional supports and classroom accommodation(s) to each student customarily provided in accordance with the student’s IEP. These instructional supports and accommodations also applied to the collection of student responses for the CAA for Science.

### Processing and Scoring

The CAA for Science was administered online only and required two internet-connected devices: a student testing device and a separate device the test examiner used to start a test session through the Test Administrator Interface. Test examiners could also use their device to open a *DFA* document, with which the test examiner guided the student through the test. The CAA for Science required the installation of CAASPP secure browsers on student testing devices. These were the same secure browsers used for the other online CAASPP assessments.

### Test Security and Confidentiality

#### ETS’ Office of Testing Integrity

The Office of Testing Integrity (OTI) is a division of ETS that provides quality assurance services for all ETS-managed testing programs. This division resides in the ETS legal department. The Office of Professional Standards Compliance at ETS publishes and maintains the *ETS Standards for Quality and Fairness* (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.

#### Procedures to Maintain Standardization of Test Security

Test security requires the accounting of all secure materials before, during, and after each test administration. The LEA CAASPP coordinator is responsible for keeping all test materials secure, keeping student information confidential, and making sure the CAASPP test site coordinators and test examiners 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 examiner is responsible for reporting testing incidents to the CAASPP test site coordinator and securely destroying printed embedded PTs (CDE, 2019e).

The following measures ensured the security of CAASPP System assessments administered in 2018–2019:

* LEA CAASPP coordinators and test site coordinators must have signed and submitted a “CAASPP Test Security Agreement for LEA CAASPP coordinators and CAASPP test site coordinators” form in TOMS before ETS granted the coordinators access to TOMS (5 *CCR*, Education, Section 859[a]).
* Anyone having access to the testing materials must have electronically signed and submitted a “Test Security Affidavit for Test Examiners, Test Administrators, Proctors, Translators, Scribes, and Any Other Person Having Access to CAASPP Tests” form electronically in TOMS before receiving access to any testing materials (5*CCR*, Section 859[c]).

In addition, it was the responsibility of every participant in the CAASPP System to report immediately any violation or suspected violation of test security or confidentiality. The CAASPP test site coordinator reported to the LEA CAASPP coordinator, and the LEA CAASPP coordinator reported to the CDE within 24 hours of the incident (5 *CCR*, Section 859[e]).

#### Security of Electronic Files Using a Firewall

A firewall software is currently used to prevent 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; and to Concord and Sacramento, California.

All electronic applications that are included in TOMS remain protected by the ETS firewall software at all times. Due to the sensitive nature of the student information processed by TOMS, the firewall plays a significant role in maintaining assurance of confidentiality among the users of this information.

Refer to section [*1.9 Systems Overview and Functionality*](#_Systems_Overview_and) in [*Chapter 1: Introduction*](#_Introduction) for more information on TOMS.

#### Transfer of Scores via Secure Data Exchange

Due to 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 only temporarily stored 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.

ETS enters information about the files posted to the SFTP server in a web form on a SharePoint website. A CDE staff member reviews this log throughout the day to check the status of deliverables and downloads and deletes the file from the SFTP server when its status shows it has 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 data is collected during operational testing. Only individuals with the appropriate credentials can access the data. ETS builds all interfaces with the most stringent security considerations, including interfaces with data encryption for databases that store test items and student data. ETS applies best and up-to-date security practices, including system-to-system authentication and authorization, in all solution designs.

All stored test content and student data is encrypted. 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.

#### Statistical Analysis on Secure Servers

During all CAASPP testing, ETS information technology staff retrieves data files from the American Institutes for Research and loads them into a database. The ETS Data Quality Services staff extracts the data from the database and performs 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 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.

#### Security and Test Administration Incident Reporting System (STAIRS) 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, compromises the reliability and validity of test results (CDE, 2019e). 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 must ensure that all test security and summative administration incidents are 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 the form was submitted, an email containing a case number and next steps was sent to the submitter (and to the LEA CAASPP coordinator, if the form is submitted by the CAASPP test site coordinator). The STAIRS case in TOMS provided the LEA CAASPP coordinator, the CDE, and the California Technical Assistance Center (CalTAC) with the opportunity to interact and communicate regarding the STAIRS process (CDE, 2019e).

The following types of STAIRS reports, as applicable to the CAAs, were also forwarded to the CDE:

* Security breach (where secure materials were exposed)
* Accidental access to a summative assessment
* Incorrect Statewide Student Identifier used (intentionally switched)
* Restoring a test that had been reset

Appeals requests were reviewed by the CDE. When a request to submit an Appeal was approved, the coordinator received a system-generated email with the Appeal type that was approved (CDE, 2019e).

##### 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 should report the incident within 24 hours, using the online 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 online 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 and escalation to CalTAC (for social media breaches) or the CDE (for all other breaches) 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. 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. A breach incident must be reported to the LEA CAASPP coordinator immediately.

#### Appeals

For test security incidents reported in STAIRS that resulted in a need to reset, reopen, invalidate, or restore individual online student assessments, the request was approved by the CDE. 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 are available for retrieval and review by the appropriate credentialed users within a given organization. However, the view of Appeals is restricted according to the user role as established in TOMS. 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, 2019e).

Types of appeals available during the 2018–2019 CAASPP administration are described in table 5.4.

Table 5.4 Types of Appeals in CAASPP Testing

|  |  |
| --- | --- |
| Type of Appeal | Description |
| Reset | Resetting a student’s summative assessment removes that assessment from the system and enables the student to start a new assessment from the beginning. |
| Invalidation | Invalidated summative assessments will be scored, and scores will be provided on the Student Score Report with a note that an irregularity occurred. The student(s) will be counted as participating in the calculation of the school’s participation rate for accountability purposes. |
| Re-open | Reopening a summative assessment allows a student to access an assessment that has already been submitted. |
| Restore | Restoring a summative assessment returns an assessment from the Reset status to its prior status. This action can only be performed on assessments that have been previously reset. |

### References

California Department of Education. (2018b). *Technical specifications and configuration guide for CAASPP online testing*. Sacramento, CA: California Department of Education.

California Department of Education. (2018a). *TOMS pre-administration guide for CAASPP testing*. Sacramento, CA: California Department of Education.

California Department of Education. (2019d). *2018–19 CAA for Science administration planning guide: Grade eight.* Sacramento, CA: California Department of Education.

California Department of Education. (2019b). *California Alternate Assessment for Science directions for administration, training performance task, fossils and chemical changes.* Sacramento, CA: California Department of Education.

California Department of Education. (2019a). *2018–19 California Alternate Assessment for Science test examiner tutorial*. Sacramento, CA: California Department of Education.

California Department of Education. (2019c). *CAASPP online test administration manual, 2018–19 administration.* Sacramento, CA: California Department of Education.

California Department of Education. (2019e). *Security incidents and appeals procedure guide, 2018–19 administration.* Sacramento, CA: California Department of Education.

California Department of Education. (2020). *Administering the California Alternate Assessment for Science*. Sacramento, CA: California Department of Education.

California Science Teachers Association. (2000–2019). *CA Alternate Assessment - CAA- Science*.

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

## Scoring and Reporting

Student scores for the field test of the California Alternate Assessment (CAA) for Science, given during the 2018–2019 California Assessment of Student Performance and Progress (CAASPP) administration, were not reported using CAASPP Student Score Reports. However, the percent-correct scores and preliminary indicator categories were calculated to provide the local educational agencies (LEAs) with information on student performance on the assessment. This chapter describes how the student responses were scored to determine each student’s percent-correct score and preliminary indictor category.

### CAA for Science Scoring Process

Each student was administered three embedded performance tasks (PTs), each consisting of 10 items, for a total of 12 points. Two items in each PT are worth two points.

During administration, the student’s answer to each item was entered into the CAASPP test delivery system (TDS). Instructions detailing how to administer the tests were provided by Educational Testing Service (ETS) in the secure *Embedded Performance Task Directions for Administration.* Refer to the nonsecure training test *DFA* for the type of information and instructions that were available to test examiners (California Department of Education [CDE], 2019).

Student responses to items were captured and scored in the TDS and then the data was passed directly from the quality monitoring system to the database of record to be transmitted to ETS. The percent correct and preliminary indicator were assigned to each student by ETS’ Enterprise Score Key Management System.

Although there was no formal Student Score Report for the CAA for Science field test, the LEAs were responsible for reporting the student performance results to the student’s parents/guardians. To that end, ETS provided the CDE with an aggregate file that included the mean percent-correct scores and the percentage of students scoring at each preliminary indicator category at the school, LEA, and state levels.

### Types of Scores

To provide a broad and early indication about an LEA’s implementation of the California Next Generation Science Standards Core Content Connectors (Science Connectors) on the CAA for Science, two types of scores were calculated: the percent-correct score that indicates percentage of maximum points earned by a student; and a preliminary indicator category that indicates low, medium, or high performance (implying limited, moderate, or considerable understanding of the content tested).

#### Percent Correct

The percent-correct scores are calculated for the field test items. The percent correct is calculated using the formula presented in equation 6.1. *Refer to the* [*Alternative Text for Equation 6.1*](#_Alternative_Text_for_10) *for a description of this equation*.

 (6.1)

If the student did not respond to at least one item for the embedded PT, a score of 0 (zero) was assigned for that embedded PT.

#### Preliminary Indicator Categories

The preliminary indicators are descriptive statements with corresponding threshold scores used in reporting the CAA for Science results. Indicators are considered preliminary because they are available to parents/guardians and the public before the completion of the science assessments’ development (CDE, 2018a).

There were three preliminary indicator categories to indicate high (category 3), medium (category 2), or low (category 1) performance. A student’s preliminary indicator category provided a general indication of the student’s understanding of the Science Connectors. Table 6.1 provides the description of each indicator category.

Table 6.1 Indicator Categories

|  |  |
| --- | --- |
| Category | Explanation |
| 3 | Student performance suggests a considerable understanding of the Science Connectors. |
| 2 | Student performance suggests a *moderate* understanding of the Science Connectors. |
| 1 | Student performance suggests a limited understanding of the Science Connectors. |

Students who performed at or below the chance level—the average performance expected of students responding to each item at random—were assigned to the indicator category of 1. Students who performed exceedingly well (i.e., 90 percent correct or above) were assigned the indicator category of 3. Most students are in category 2.

A group of California science educators familiar with the eligible student population reviewed and provided feedback on plans and initial drafts of preliminary indicators on December 20, 2017. The threshold scores for the three indicator categories are presented in table 6.2. Each threshold score is expressed as a percentage of the maximum possible score.

Table 6.2 Threshold Scores for Preliminary Categories

|  |  |  |
| --- | --- | --- |
| Grade Level | Required for Category 2 | Required for Category 3 |
| Grade 5 | 33% | 90% |
| Grade 8 | 33% | 90% |
| High school | 33% | 90% |

The preliminary indicator conversion table is shown for grade five in table 6.3. This table provides the percent-correct score and preliminary category for each possible raw score.

Table 6.3 Grade Five Preliminary Indicator Conversion Table

|  |  |  |
| --- | --- | --- |
| Raw Score (# of points earned) | Percent Correct | Preliminary Category |
| 0 | 0 | 1 |
| 1 | 3 | 1 |
| 2 | 6 | 1 |
| 3 | 8 | 1 |
| 4 | 11 | 1 |
| 5 | 14 | 1 |
| 6 | 17 | 1 |
| 7 | 19 | 1 |
| 8 | 22 | 1 |
| 9 | 25 | 1 |
| 10 | 28 | 1 |
| 11 | 31 | 1 |
| 12 | 33 | 2 |
| 13 | 36 | 2 |
| 14 | 39 | 2 |
| 15 | 42 | 2 |
| 16 | 44 | 2 |
| 17 | 47 | 2 |
| 18 | 50 | 2 |
| 19 | 53 | 2 |
| 20 | 56 | 2 |
| 21 | 58 | 2 |
| 22 | 61 | 2 |
| 23 | 64 | 2 |
| 24 | 67 | 2 |
| 25 | 69 | 2 |
| 26 | 72 | 2 |
| 27 | 75 | 2 |
| 28 | 78 | 2 |
| 29 | 81 | 2 |
| 30 | 83 | 2 |
| 31 | 86 | 2 |
| 32 | 89 | 2 |
| 33 | 92 | 3 |
| 34 | 94 | 3 |
| 35 | 97 | 3 |
| 36 | 100 | 3 |

The preliminary indicator conversion table is shown for grade eight in table 6.4. This table provides the percent-correct score and preliminary category for each possible raw score.

Table 6.4 Grade Eight Preliminary Indicator Conversion Table

|  |  |  |
| --- | --- | --- |
| Raw Score (# of points earned) | Percent Correct | Preliminary Category |
| 0 | 0 | 1 |
| 1 | 3 | 1 |
| 2 | 6 | 1 |
| 3 | 8 | 1 |
| 4 | 11 | 1 |
| 5 | 14 | 1 |
| 6 | 17 | 1 |
| 7 | 19 | 1 |
| 8 | 22 | 1 |
| 9 | 25 | 1 |
| 10 | 28 | 1 |
| 11 | 31 | 1 |
| 12 | 33 | 2 |
| 13 | 36 | 2 |
| 14 | 39 | 2 |
| 15 | 42 | 2 |
| 16 | 44 | 2 |
| 17 | 47 | 2 |
| 18 | 50 | 2 |
| 19 | 53 | 2 |
| 20 | 56 | 2 |
| 21 | 58 | 2 |
| 22 | 61 | 2 |
| 23 | 64 | 2 |
| 24 | 67 | 2 |
| 25 | 69 | 2 |
| 26 | 72 | 2 |
| 27 | 75 | 2 |
| 28 | 78 | 2 |
| 29 | 81 | 2 |
| 30 | 83 | 2 |
| 31 | 86 | 2 |
| 32 | 89 | 2 |
| 33 | 92 | 3 |
| 34 | 94 | 3 |
| 35 | 97 | 3 |
| 36 | 100 | 3 |

The preliminary indicator conversion table is shown for high school in table 6.5. This table provides the percent-correct score and preliminary category for each possible raw score.

Table 6.5 High School Preliminary Indicator Conversion Table

|  |  |  |
| --- | --- | --- |
| Raw Score (# of points earned) | Percent Correct | Preliminary Category |
| 0 | 0 | 1 |
| 1 | 3 | 1 |
| 2 | 6 | 1 |
| 3 | 8 | 1 |
| 4 | 11 | 1 |
| 5 | 14 | 1 |
| 6 | 17 | 1 |
| 7 | 19 | 1 |
| 8 | 22 | 1 |
| 9 | 25 | 1 |
| 10 | 28 | 1 |
| 11 | 31 | 1 |
| 12 | 33 | 2 |
| 13 | 36 | 2 |
| 14 | 39 | 2 |
| 15 | 42 | 2 |
| 16 | 44 | 2 |
| 17 | 47 | 2 |
| 18 | 50 | 2 |
| 19 | 53 | 2 |
| 20 | 56 | 2 |
| 21 | 58 | 2 |
| 22 | 61 | 2 |
| 23 | 64 | 2 |
| 24 | 67 | 2 |
| 25 | 69 | 2 |
| 26 | 72 | 2 |
| 27 | 75 | 2 |
| 28 | 78 | 2 |
| 29 | 81 | 2 |
| 30 | 83 | 2 |
| 31 | 86 | 2 |
| 32 | 89 | 2 |
| 33 | 92 | 3 |
| 34 | 94 | 3 |
| 35 | 97 | 3 |
| 36 | 100 | 3 |

Table 6.A.1 shows, for several groups of students taking the grade five assessment, the mean raw score, mean percent correct, and the percentage of students at each category level. This information is provided for the total group of students and for each of several demographic student groups (e.g., gender, ethnicity, primary disability, etc.). Table 6.A.2 and table 6.A.3 show the same information for the grade eight test and the high school test, respectively.

#### Aggregate Score Reporting

To provide meaningful results to the stakeholders, test scores for a given grade are aggregated at the school, LEA, county, and state levels. (A direct funded charter school is reported as a separate LEA.) The aggregated scores are generated for selected groups of interest to the CDE (e.g., gender, ethnicity, primary disability, etc.) and for the total population.

Statistics summarizing student performance by content area and grade for the selected groups of students are provided in [appendix 6.A](#_Appendix_6.A:_Demographic). In table 6.A.1 through table 6.A.3, students are grouped by demographic characteristics, including gender, ethnicity, English language fluency, economic status (disadvantaged or not), primary disability, migrant status, and ethnicity by economic status. For each demographic group, the table shows the number of students with a valid raw score, the raw-score means and standard deviations (SDs), the percent-correct means and SDs, and the percentage of students in each preliminary indicator category.

Table 6.6 lists the demographic groups for which these statistics are reported. To protect students’ privacy, when the number of students in a student group is 10 or fewer, the summary statistics are not reported and are replaced in the table by “N/A.”

Table 6.6 Demographic Student Groups to Be Reported

|  |  |
| --- | --- |
| Category | Student Groups |
| Gender | * Male * Female |
| 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 |
| English-Language Fluency | * English only * Initial fluent English proficient * English learner * Reclassified fluent English proficient * To be determined * English proficiency unknown |
| Economic Status | * Not economically disadvantaged * Economically disadvantaged |
| Primary Disability Type | * Intellectual disability * Hearing impairment * Speech or language impairment * Visual impairment * Emotional disturbance * Orthopedic impairment * Other health impairment * Specific learning disability * Deaf-blindness * Multiple disabilities * Autism * Traumatic brain injury * Not classified[[5]](#footnote-6) |
| Migrant Status | * Eligible for the Title I Part C Migrant Program (Migrant) * Not eligible for the Title I Part C Migrant Program (Nonmigrant) |

### Survey Questions Regarding Test Administration

Survey questions were presented at the end of each embedded PT. The test examiner entered responses to these questions into the TDS.

#### Student Engagement

In the survey that follows, the test examiner answered the question regarding how engaged the student was when completing the embedded PT. The survey question was administered at the end of each embedded PT. An example of an engagement question is as follows:

1. How engaged was your student with this test you just administered?

* A—Fully engaged
* B—Moderately engaged
* C—Minimally engaged

The summary of the data results is provided in subsection [*8.1.1 Student Survey*](#_Student_Survey).

#### Individualization of the Test

The CAA for Science is designed to strike a careful balance between standardized administration and maximizing student engagement. To meet this goal, some parts of each embedded PT can be individualized to improve student engagement.

For the field test administration, test examiners were instructed to review the activities associated with each embedded PT and decide whether the exemplar activity met a student’s needs or if an individualized activity was appropriate. The test examiner documented the use of individualizations via survey.

Examples of the text of the individualization questions are as follows:

1. Did you individualize any aspect of Orienting Activity #1 and the first five questions, where permitted?

* Yes
* No
* If yes, and you used specific materials, briefly describe:

1. Did you individualize any aspect of Orienting Activity #2 and the last five questions, where permitted?

* Yes
* No
* If yes, and you used specific materials, briefly describe:

The summary of the individualization is provided in subsection[*5.5 Accessibility Features for the Field Test*](#_Accessibility_Features_for).

### References

California Department of Education. (2019). *California Alternate Assessment for Science directions for administration, training performance task, fossils and chemical changes.* Sacramento, CA: California Department of Education.

California Department of Education. (2018). *Science assessments preliminary Indicators FAQ, question 7.*

### Accessibility Information

#### Alternative Text for Equation 6.1

Percent correct equals the number of points earned for all items divided by the maximum number of points for all items. *(Return to* [*equation 6.1*](#equation_6_1)*.)*

### Appendix 6.A: Demographic Summaries

Table 6.A.1 Demographic Summary for Grade Five

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Student Group | Number Tested | Mean Raw Score | SD of Raw Scores | Percent in Preliminary Category 1 | Percent in Preliminary Category 2 | Percent in Preliminary Category 3 |
| All students with valid scores | 5,131 | 21 | 10 | 16% | 77% | 7% |
| Male | 3,432 | 21 | 10 | 17% | 76% | 7% |
| Female | 1,699 | 21 | 10 | 15% | 80% | 5% |
| American Indian or Alaska Native | 27 | 25 | 7 | 4% | 78% | 19% |
| Asian | 409 | 19 | 10 | 21% | 74% | 5% |
| Native Hawaiian or Other Pacific Islander | 24 | 20 | 9 | 17% | 83% | 0% |
| Filipino | 122 | 18 | 10 | 20% | 77% | 3% |
| Hispanic or Latino | 2,962 | 21 | 10 | 15% | 78% | 7% |
| Black or African American | 419 | 21 | 10 | 16% | 78% | 5% |
| White | 943 | 21 | 10 | 17% | 75% | 7% |
| Two or more races | 185 | 21 | 10 | 17% | 76% | 7% |
| English only | 3,038 | 21 | 10 | 16% | 77% | 7% |
| Initial fluent English proficient | 57 | 19 | 9 | 16% | 84% | 0% |
| English learner | 1,600 | 21 | 10 | 16% | 77% | 7% |
| Reclassified fluent English proficient | 430 | 21 | 10 | 15% | 80% | 6% |
| To be determined | 0 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 6 | N/A | N/A | N/A | N/A | N/A |
| Intellectual disability | 1,813 | 20 | 9 | 15% | 82% | 3% |
| Hearing impairment | 48 | 24 | 9 | 10% | 79% | 10% |
| Speech or language impairment | 121 | 28 | 5 | 1% | 83% | 16% |
| Visual impairment | 18 | 17 | 13 | 33% | 61% | 6% |
| Emotional disturbance | 23 | 30 | 4 | 0% | 78% | 22% |
| Orthopedic impairment | 196 | 16 | 12 | 35% | 64% | 1% |
| Other health impairment | 298 | 25 | 9 | 8% | 78% | 14% |
| Specific learning disability | 391 | 30 | 4 | 0% | 79% | 21% |
| Deaf-blindness | 3 | N/A | N/A | N/A | N/A | N/A |
| Multiple disabilities | 278 | 12 | 11 | 49% | 50% | 1% |
| Autism | 1,922 | 20 | 10 | 16% | 78% | 6% |
| Traumatic brain injury | 20 | 20 | 12 | 25% | 70% | 5% |
| Not classified | 0 | N/A | N/A | N/A | N/A | N/A |
| Not economically disadvantaged | 1,677 | 19 | 10 | 21% | 74% | 5% |
| Economically disadvantaged | 3,454 | 22 | 9 | 14% | 79% | 7% |
| Migrant | 34 | 25 | 7 | 6% | 85% | 9% |
| Nonmigrant | 5,097 | 21 | 10 | 16% | 77% | 7% |
| American Indian or Alaska Native—Not economically disadvantaged | 11 | 24 | 7 | 0% | 82% | 18% |
| American Indian or Alaska Native—Economically disadvantaged | 16 | 26 | 8 | 6% | 75% | 19% |
| Asian—Not economically disadvantaged | 231 | 18 | 10 | 25% | 71% | 3% |
| Asian—Economically disadvantaged | 178 | 21 | 10 | 16% | 77% | 7% |
| Native Hawaiian or Other Pacific Islander—Not economically disadvantaged | 8 | N/A | N/A | N/A | N/A | N/A |
| Native Hawaiian or Other Pacific Islander—Economically disadvantaged | 16 | 23 | 5 | 0% | 100% | 0% |
| Filipino—Not economically disadvantaged | 82 | 18 | 10 | 20% | 79% | 1% |
| Filipino—Economically disadvantaged | 40 | 18 | 10 | 20% | 73% | 8% |
| Hispanic or Latino—Not economically disadvantaged | 604 | 19 | 10 | 20% | 75% | 4% |
| Hispanic or Latino—Economically disadvantaged | 2,358 | 22 | 9 | 13% | 79% | 7% |
| Black or African American—Not economically disadvantaged | 119 | 19 | 10 | 22% | 75% | 3% |
| Black or African American—Economically disadvantaged | 300 | 21 | 10 | 14% | 80% | 6% |
| White—Not economically disadvantaged | 512 | 19 | 10 | 21% | 73% | 6% |
| White—Economically disadvantaged | 431 | 22 | 10 | 13% | 78% | 9% |
| Two or more races—Not economically disadvantaged | 94 | 20 | 11 | 19% | 73% | 7% |
| Two or more races—Economically disadvantaged | 91 | 22 | 9 | 14% | 79% | 7% |

Table 6.A.2 Demographic Summary for Grade Eight

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Student Group | Number Tested | Mean Raw Score | SD of Raw Scores | Percent in Preliminary Category 1 | Percent in Preliminary Category 2 | Percent in Preliminary Category 3 |
| All students with valid scores | 5,217 | 20 | 10 | 16% | 77% | 7% |
| Male | 3,501 | 21 | 10 | 15% | 76% | 8% |
| Female | 1,716 | 20 | 9 | 16% | 79% | 6% |
| American Indian or Alaska Native | 29 | 24 | 8 | 3% | 83% | 14% |
| Asian | 398 | 17 | 10 | 23% | 73% | 4% |
| Native Hawaiian or Other Pacific Islander | 29 | 17 | 10 | 24% | 72% | 3% |
| Filipino | 138 | 19 | 9 | 15% | 82% | 3% |
| Hispanic or Latino | 2,979 | 21 | 10 | 15% | 78% | 7% |
| Black or African American | 428 | 21 | 10 | 15% | 76% | 9% |
| White | 1,014 | 21 | 10 | 15% | 75% | 9% |
| Two or more races | 165 | 20 | 9 | 14% | 81% | 5% |
| English only | 2,983 | 20 | 10 | 16% | 76% | 8% |
| Initial fluent English proficient | 81 | 16 | 9 | 25% | 75% | 0% |
| English learner | 1,454 | 20 | 10 | 16% | 77% | 6% |
| Reclassified fluent English proficient | 696 | 21 | 9 | 12% | 80% | 8% |
| 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 |
| Intellectual disability | 2,108 | 19 | 9 | 15% | 81% | 4% |
| Hearing impairment | 50 | 25 | 7 | 4% | 88% | 8% |
| Speech or language impairment | 83 | 26 | 6 | 2% | 83% | 14% |
| Visual impairment | 23 | 15 | 11 | 35% | 61% | 4% |
| Emotional disturbance | 29 | 29 | 4 | 0% | 72% | 28% |
| Orthopedic impairment | 215 | 15 | 12 | 37% | 60% | 3% |
| Other health impairment | 279 | 25 | 8 | 6% | 83% | 11% |
| Specific learning disability | 375 | 29 | 4 | 0% | 76% | 24% |
| Deaf-blindness | 0 | N/A | N/A | N/A | N/A | N/A |
| Multiple disabilities | 257 | 11 | 11 | 49% | 49% | 2% |
| Autism | 1,771 | 20 | 9 | 15% | 78% | 7% |
| Traumatic brain injury | 27 | 21 | 12 | 22% | 67% | 11% |
| Not classified | 0 | N/A | N/A | N/A | N/A | N/A |
| Not economically disadvantaged | 1,857 | 19 | 10 | 19% | 75% | 5% |
| Economically disadvantaged | 3,360 | 21 | 9 | 13% | 78% | 8% |
| Migrant | 42 | 23 | 9 | 10% | 79% | 12% |
| Nonmigrant | 5,175 | 20 | 10 | 16% | 77% | 7% |
| American Indian or Alaska Native—Not economically disadvantaged | 8 | N/A | N/A | N/A | N/A | N/A |
| American Indian or Alaska Native—Economically disadvantaged | 21 | 26 | 7 | 0% | 81% | 19% |
| Asian—Not economically disadvantaged | 239 | 17 | 10 | 25% | 73% | 2% |
| Asian—Economically disadvantaged | 159 | 18 | 10 | 21% | 73% | 6% |
| Native Hawaiian or Other Pacific Islander—Not economically disadvantaged | 8 | N/A | N/A | N/A | N/A | N/A |
| Native Hawaiian or Other Pacific Islander—Economically disadvantaged | 21 | 17 | 9 | 24% | 76% | 0% |
| Filipino—Not economically disadvantaged | 95 | 17 | 9 | 20% | 77% | 3% |
| Filipino—Economically disadvantaged | 43 | 22 | 7 | 5% | 93% | 2% |
| Hispanic or Latino—Not economically disadvantaged | 681 | 19 | 10 | 20% | 75% | 5% |
| Hispanic or Latino—Economically disadvantaged | 2,298 | 21 | 9 | 13% | 79% | 8% |
| Black or African American—Not economically disadvantaged | 126 | 19 | 10 | 20% | 73% | 7% |
| Black or African American—Economically disadvantaged | 302 | 21 | 9 | 14% | 77% | 10% |
| White—Not economically disadvantaged | 591 | 20 | 10 | 16% | 76% | 7% |
| White—Economically disadvantaged | 423 | 22 | 10 | 14% | 74% | 12% |
| Two or more races—Not economically disadvantaged | 97 | 19 | 9 | 16% | 79% | 4% |
| Two or more races—Economically disadvantaged | 68 | 22 | 9 | 10% | 82% | 7% |

Table 6.A.3 Demographic Summary for High School

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Student Group | Number Tested | Mean Raw Score | SD of Raw Scores | Percent in Preliminary Category 1 | Percent in Preliminary Category 2 | Percent in Preliminary Category 3 |
| All students with valid scores | 8,814 | 19 | 10 | 20% | 73% | 6% |
| Male | 5,832 | 19 | 10 | 21% | 72% | 7% |
| Female | 2,982 | 19 | 10 | 20% | 75% | 5% |
| American Indian or Alaska Native | 55 | 23 | 10 | 11% | 73% | 16% |
| Asian | 757 | 17 | 10 | 26% | 71% | 3% |
| Native Hawaiian or Other Pacific Islander | 49 | 18 | 9 | 18% | 82% | 0% |
| Filipino | 308 | 16 | 10 | 27% | 70% | 3% |
| Hispanic or Latino | 4,602 | 19 | 10 | 20% | 75% | 6% |
| Black or African American | 710 | 19 | 10 | 20% | 74% | 6% |
| White | 2,038 | 20 | 11 | 21% | 70% | 9% |
| Two or more races | 240 | 21 | 10 | 14% | 78% | 9% |
| English only | 5,291 | 19 | 11 | 21% | 72% | 7% |
| Initial fluent English proficient | 154 | 17 | 10 | 23% | 71% | 5% |
| English learner | 2,049 | 18 | 10 | 23% | 73% | 4% |
| Reclassified fluent English proficient | 1,304 | 20 | 9 | 15% | 80% | 5% |
| To be determined | 6 | N/A | N/A | N/A | N/A | N/A |
| English proficiency unknown | 10 | N/A | N/A | N/A | N/A | N/A |
| Intellectual disability | 3,893 | 19 | 9 | 17% | 79% | 4% |
| Hearing impairment | 74 | 21 | 9 | 12% | 84% | 4% |
| Speech or language impairment | 65 | 27 | 6 | 3% | 85% | 12% |
| Visual impairment | 57 | 12 | 12 | 53% | 39% | 9% |
| Emotional disturbance | 55 | 27 | 9 | 9% | 58% | 33% |
| Orthopedic impairment | 441 | 16 | 12 | 36% | 60% | 4% |
| Other health impairment | 330 | 25 | 8 | 7% | 78% | 15% |
| Specific learning disability | 401 | 27 | 8 | 6% | 70% | 24% |
| Deaf-blindness | 3 | N/A | N/A | N/A | N/A | N/A |
| Multiple disabilities | 532 | 10 | 11 | 59% | 39% | 2% |
| Autism | 2,895 | 19 | 10 | 19% | 74% | 7% |
| Traumatic brain injury | 68 | 21 | 11 | 21% | 71% | 9% |
| Not classified | 0 | N/A | N/A | N/A | N/A | N/A |
| Not economically disadvantaged | 3,550 | 18 | 11 | 25% | 70% | 5% |
| Economically disadvantaged | 5,264 | 20 | 10 | 18% | 76% | 7% |
| Migrant | 28 | 20 | 10 | 18% | 71% | 11% |
| Nonmigrant | 8,786 | 19 | 10 | 20% | 73% | 6% |
| American Indian or Alaska Native—Not economically disadvantaged | 19 | 22 | 11 | 16% | 74% | 11% |
| American Indian or Alaska Native—Economically disadvantaged | 36 | 24 | 9 | 8% | 72% | 19% |
| Asian—Not economically disadvantaged | 453 | 16 | 10 | 26% | 73% | 2% |
| Asian—Economically disadvantaged | 304 | 17 | 11 | 26% | 69% | 6% |
| Native Hawaiian or Other Pacific Islander—Not economically disadvantaged | 20 | 18 | 10 | 20% | 80% | 0% |
| Native Hawaiian or Other Pacific Islander—Economically disadvantaged | 29 | 19 | 9 | 17% | 83% | 0% |
| Filipino—Not economically disadvantaged | 193 | 16 | 10 | 28% | 69% | 3% |
| Filipino—Economically disadvantaged | 115 | 16 | 10 | 26% | 71% | 3% |
| Hispanic or Latino—Not economically disadvantaged | 1,202 | 18 | 11 | 24% | 70% | 5% |
| Hispanic or Latino—Economically disadvantaged | 3,400 | 20 | 10 | 18% | 77% | 6% |
| Black or African American—Not economically disadvantaged | 252 | 18 | 11 | 26% | 69% | 5% |
| Black or African American—Economically disadvantaged | 458 | 20 | 10 | 17% | 76% | 7% |
| White—Not economically disadvantaged | 1,242 | 18 | 11 | 25% | 67% | 7% |
| White—Economically disadvantaged | 796 | 22 | 10 | 13% | 75% | 12% |
| Two or more races—Not economically disadvantaged | 144 | 20 | 10 | 16% | 80% | 4% |
| Two or more races—Economically disadvantaged | 96 | 22 | 10 | 10% | 74% | 16% |

## Psychometric Analyses

This chapter summarizes the results of the psychometric analyses conducted for the California Alternate Assessment (CAA) for Science field test administered during the 2018–‍2019 California Assessment of Student Performance and Progress (CAASPP) administration.

### Overview

This chapter describes the data samples used for the statistical analyses, presents the results of the item and test analyses, and provides explanations for all statistical procedures implemented in the psychometric analyses. Those procedures include item analyses, DIF analyses, item response theory (IRT) calibration, computation of reliability, and standard errors of measurement. The procedures designed to ensure the validity of score uses and interpretations also are provided.

#### Summary of the Analyses

Educational Testing Service (ETS) conducted the following analyses for the CAA for Science. Each analysis is described in the text of this chapter, and the results are provided in the appendices.

1. **Classical Item Analyses—**Classical item analysis for the CAA for Science is discussed in subsection [*7.2 Classical Item Analysis Statistics*](#_Classical_Item_Analysis). [Appendix 7.A](#_Appendix_7.A:_Classical) presents results of the classical item analyses, including item difficulty indices, item-total correlation, and the distribution of item scores for each item. The item type and item flags are also provided.
2. **IRT Analyses—**IRT calibration analyses for the CAA for Science are described in subsection [*7.6 Item Response Theory (IRT) Analyses*](#_Item_Response_Theory). [Appendix 7.E](#_Appendix_7.E:_Item) includes the item difficulty parameter estimates (*b*-value) for all of the items in each test. For polytomous items, step difficulty estimates (*d*-parameters) are also provided.
3. **Omission and Completion Analyses—**The omit rate and item difficulty information for the CAA for Science are described in subsection [*7.3 Omission and Completion Rates*](#_Omission_and_Completion), the item difficulty and omit rates are provided in table 7.E.5 through table 7.E.7, and the results of the completion analyses are presented in [appendix 7.B](#_Appendix_7.B:_Completion). These analyses show how many students completed each of the embedded PTs.
4. **DIF Analyses—**DIF analysis for the CAA for Science is described in subsection [*7.5 DIF Analyses*](#_DIF_Analyses). [Appendix 7.D](#_Appendix_7.D:_DIF) presents the results of the DIF analyses for all items with sufficient student sample sizes. The distributions of items across DIF categories are included.
5. **Reliability Analyses—**Reliability estimation for the CAA for Science is illustrated in subsection [*7.7 Reliability Analyses*](#_Reliability_Analyses). Table 7.F.1 through table 7.F.6 in [appendix 7.F](#_Appendix_7.F:_Reliability) provide results of the reliability analyses of total test scores for the population as a whole, selected student groups of interest (e.g., gender, ethnicity, etc.), and each version.
6. **Validity Evidence—**Validity evidence related to the CAA for Science is discussed in subsection [*7.8 Validity Evidence*](#_Validity_Evidence). Table 7.G.2 through table 7.G.7 provide summary statistics regarding the students’ level of engagement with each embedded PT. Table 7.G.9 through table 7.G.14 in [appendix 7.G](#_Appendix_7.G:_Validity) present distributions of the observed testing time to complete the total test for each quartile group.

#### Sample for the Analyses

In general, analyses included in the technical report are based on all students in the tested population with valid scores available at the time of analysis. The actual data sample used depends on both the time the data became available as well as the information (e.g., student demographic information, scores for each embedded performance task [PT], etc.) contained in that data at the time of the analyses.

For the 2018–2019 CAA for Science, a small number of student scores were excluded from the final production data as a result of the data validation process. Students who did not answer at least one item for each of the three embedded PTs were excluded from the analysis sample for the classical item analysis, DIF analyses, and IRT calibrations.

Table 7.1 provides the number of students in each of the grade levels for the data analyses. The data in table 7.1 reveals that the majority of students in each grade level completed the CAA for Science. Grade twelve had a high percentage of students who did not take the CAA for Science—2,193 students out of 8,141 students, or 26.9 percent. This high rate of noncompleters can be partially attributed to students over the age of 18 who were registered to take the CAA for Science but did not do so.

Table 7.1. Analysis Data Sources

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade Level | Analysis Sample, Completers | Final Data, Completers | Final Data, Partial Completers | Final Data, Noncompleters | Final Data, Total Number Tested |
| Grade 5 | 4,414 | 5,131 | 150 | 566 | 5,847 |
| Grade 8 | 4,520 | 5,217 | 129 | 654 | 6,000 |
| High school—All grades | 7,264 | 8,820 | 240 | 3,073 | 12,133 |
| Grade 10 | 317 | 377 | 8 | 136 | 521 |
| Grade 11 | 2,289 | 2,660 | 67 | 744 | 3,471 |
| Grade 12 | 4,658 | 5,783 | 165 | 2,193 | 8,141 |

### Classical Item Analysis Statistics

Classical item analyses were used to evaluate the field test items with respect to item difficulty, item discrimination, and student performance on the embedded PT items.

The classical item analyses include the computation of item difficulty indices and item-total correlations. Flagging rules based on these statistics identified items not performing as expected. The omit rate of each item and the distribution of scores on each polytomous item are also included in the classical item analyses.

#### Classical Item Difficulty Indices (*p*-value and Average Item Score)

For dichotomous items, item difficulty is indicated by the *p*-value, which is the proportion of students who answer an item correctly. The range of possible *p*-values is from 0.00 to 1.00. Items with higher *p*-values are easier items; those with lower *p*-values are more difficult items. Dichotomous items are flagged for review if their *p*-values are above 0.95 (i.e., too easy) or below 0.33 (i.e., too difficult).

The formula for *p*-value for a dichotomous item is presented in equation 7.1. *Refer to the* [*Alternative Text for Equation 7.1*](#_Alternative_Text_for) *for a description of this equation.*

 (7.1)

where,

*Xij* is the score (1 or 0) 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, difficulty is indicated by the average item score (AIS). 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 30 percent to 80 percent of the maximum obtainable item score; items with values outside this range are flagged for review. To facilitate interpretation, the AIS values for polytomous items are often expressed as a proportion of the maximum possible score, which is analogous to the *p-*values of dichotomous items.

For polytomous items, the *p-value* is defined as presented in equation 7.2. *Refer to the* [*Alternative Text for Equation 7.2*](#_Alternative_Text_for_1) *for a description of this equation.*

 (7.2)

where,

 is the score received for a given polytomous item *i* for student *j*,

 is the total number of students who were presented with item *i*, and

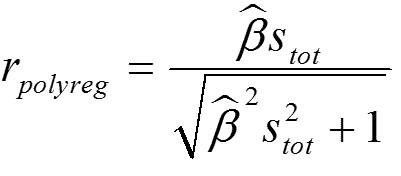
*Max (Xi)* is the maximum score on item *i*.

#### Item Discrimination (Item-Total Correlation)

An item-total correlation describes the relationship between students’ performance on a specific item and their performance on the total test.

In general, the possible range of the item-total correlation is from -1.0 (for a perfect negative relationship) to 1.0 (for a perfect positive relationship). A relatively high positive item-total correlation is desired, as it indicates that students with higher scores on the assessment tended to perform better on the item than students with lower test scores. A negative item‑total correlation, which indicates that students with low scores on the assessment are more likely to get higher scores on the item than students with high scores on the assessment, typically signifies a problem with the item.

Because the product-moment correlation is limited by the distributions of the variables being correlated, the item discrimination index used in these analyses is a variation of the biserial correlation for dichotomous items or the polyserial correlation for polytomous items. This statistic is an estimate of the correlation between the criterion and an unobservable continuous variable assumed to determine performance on the item. The criterion is, in this case, the student’s total raw score from the three embedded PTs. The estimation formula is presented in equation 7.3. *Refer to the* [*Alternative Text for Equation 7.3*](#_Alternative_Text_for_2) *for a description of this equation.*

 (7.3)

where,

 is the estimated slope of the linear regression of the unobservable continuous variable (assumed to account for the item response) on the criterion, and

 is the standard deviation (SD) of the criterion (the students’ total raw 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 *k* possible score values, there are *k*-1 regressions. Beta (*β*) is the common slope for all *k*-1 regressions.

Desired values for this correlation are positive and larger than 0.20. Negative item-total correlations indicate that low-ability students tend to obtain higher scores on the item than high-ability students, an indication that the scoring key may be incorrect, or the item did not function as intended for the students taking the CAA for Science. Therefore, items with item-total correlations below 0.20 are flagged for review.

#### Distribution of Item Scores

For polytomous items, examination of the distribution of scores helps to show how well the items performed. If no students receive the highest possible score, the item may not be functioning as expected. The item may be confusing, poorly worded, or just unexpectedly difficult; the scoring rubric may be flawed; or students may not have had the opportunity to learn the content tested by the item. If all or most students score at the extreme ends of the distribution—that is, students receive either full credit or zero credit, but no partial credit—there may be problems with the item or the rubric.

Items with a low percentage (i.e., less than 3 percent) of students obtaining any possible item score are flagged for further review. Such items may pose problems during the IRT calibrations. They need to be carefully reviewed and may need to be excluded from the item calibration analyses.

#### Summary of Classical Item Analysis Flagging Criteria

Items are flagged for review if the item analysis yields any of the following results:

1. The *p-*value is above 0.95 for dichotomous items or above 0.80 for polytomous items.
2. The *p-*value is below 0.33 for dichotomous items or below 0.30 for polytomous items.
3. Item-total correlation (polyserial) is below 0.20.
4. Among the highest-performing students (the top 20 percent), the number of students choosing any distractor is greater than the number of those choosing the key.
5. The omit rate is above 5 percent for dichotomous items or above 15 percent for polytomous items.
6. Any of the possible scores on a polytomous item is earned by less than 3 percent of the students.

Refer to Note 1 of [appendix 7.A](#_Appendix_7.A:_Classical) for the flagging symbols, descriptions, and their criteria.

ETS’ psychometric staff and content assessment development staff reviewed each of the flagged items and summarized the classical item results for the California Department of Education (CDE), with recommendations for subsequent analyses of the items. The classical item statistics were entered into the item bank for use by the assessment development team for test assembly for future operational administrations.

#### Classical Item Analysis Results Summary

This subsection presents tables of the classical item analysis results for the 2018–2019 field test items. Table 7.2 presents the *p-*value—calculated when the AIS for the polytomous items was divided by the maximum possible score for that item transforming the AIS to a *p*‑value—and item-total correlation information by test.

Table 7.2 Classical Item Statistics

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level and Version | Number of Students | Number of Items | Number of Points | Mean *p-*value | Minimum *p-*value | Maximum *p-*value | Mean Item-Total Correlation |
| Grade 5 Version 1 | 2,290 | 30 | 36 | 0.67 | 0.38 | 0.92 | 0.55 |
| Grade 5 Version 2 | 2,123 | 30 | 36 | 0.67 | 0.35 | 0.91 | 0.58 |
| Grade 8 Version 1 | 2,357 | 30 | 36 | 0.66 | 0.39 | 0.93 | 0.58 |
| Grade 8 Version 2 | 2,163 | 30 | 36 | 0.64 | 0.29 | 0.93 | 0.56 |
| High school Version 1 | 3,836 | 30 | 36 | 0.65 | 0.33 | 0.90 | 0.55 |
| High school Version 2 | 3,428 | 30 | 36 | 0.64 | 0.33 | 0.90 | 0.56 |

There is a range of item difficulties with the *p-*values ranging from 0.29 to 0.93 with the mean *p-*values ranging from 0.64 to 0.67. The CAA for Science had a wide range of item difficulties, with some items being easy for the students (items with *p-*values around 0.90) and some items being difficult for the students (items with *p-*values below 0.33). However, in general, the items were about average difficulty for the students. The mean item-total correlations range from 0.55 to 0.58. These values of the item-total correlations indicate that the items have sufficient levels of discrimination.

Detailed results of the item analyses for each item by grade are presented in [appendix 7.A](#_Appendix_7.A:_Classical). The classical item analyses in [appendix 7.A](#_Appendix_7.A:_Classical) are based on the contents of the data file available in June 2018. Table 7.A.1 provides the summary statistics for each embedded PT. The item statistics, including AIS, *p-*value, item-total correlation, statistical flagging criteria, and item type are listed in those tables. The distribution of item scores on each item is presented in table 7.A.5 through table 7.A.7.

### Omission and Completion Rates

#### Omit Rates

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. An item is considered “omitted” when the item has been presented to the student but has not been answered (i.e., left blank) in the middle of an administered assessment wherein the student has been presented with, and responded to, successive items.

Table 7.A.2 through table 7.A.7 in [appendix 7.A](#_Appendix_7.A:_Classical) include the item omit rates. The items with high omit rates were flagged. The omit rate threshold for flagging individual items was 5 percent for selected-response items and 15 percent for polytomous items. Only one item out of 124 items was flagged for a high omit rate. An omit response was scored as zero and included in the N-count for that item (i.e., the number of students who answered the item) when calculating item statistics. An item that was omitted by design was not scored, and the student was not included in the N-count for the item.

Table 7.E.5 through table 7.E.7 present the omit rate and IRT item difficulty for each item.

#### Completion Rates

Completion rates indicate the proportion of students who completed each of the three embedded PTs on the test. A student’s record for the CAA for Science is not considered complete unless the student answered at least one test question from each of the three embedded PTs. The completion rates are presented in [appendix 7.B](#_Appendix_7.B:_Completion). Data used in [appendix 7.B](#_Appendix_7.B:_Completion) is based on all registered students in the full student population.

Table 7.B.1 provides the percentage of registered students who completed none, one, two, or three of the embedded PTs. Less than 3 percent of the students partially completed the CAA for Science PTs by completing only one or two of the three PTs. At grades five and eight, approximately 10 percent of the students completed none of the PTs while over 87 percent of the students at these two grades completed all three PTs. The percentage of students partially completing or not completing the CAA for Science are similar to the percentage of partial completers and noncompleters for the CAAs for English language arts/‌literacy (ELA) and mathematics assessments.

For high school, over 20 percent of the registered students completed none of the PTs and just over 70 percent of the students completed all three PTs. The percentage of high school students not completing any of the CAA for Science PTs is higher than the CAAs for ELA and mathematics noncompletion rates for grade eleven. The decrease in test-taking at the high school level, and grade twelve in particular, can be partially attributed to students who were over the age of 18 and who were registered to take the CAA for Science, but then do not attempt any of the embedded PTs.

Table 7.B.2 presents the percentage of students who completed each embedded PT. For grades five and eight, over 87 percent of the students completed all three embedded PTs. For high school, the percentage of students completing the embedded PTs was between 73 percent and 74 percent.

In general, at the high school level, grade eleven had the highest percentages of students completing the embedded PTs and grade twelve had the lowest percentages. The lowest percentage of students completing an embedded PT was 71 percent, for the high school Earth and Space Sciences embedded PT in grade twelve. The highest percentage of students completing an embedded PT was 78 percent, for the high school Physical Sciences embedded PT in grade eleven. For high school, the percentage of students completing the three embedded PTs is lower due to students who were registered to take the assessment but then were not logged on to take it.

### Task Difficulty (Overall and by Embedded PT)

The mean raw scores are provided in table 7.3 and are based on the item analysis sample. The distributions of raw scores for each test and each embedded PT are provided in [appendix 7.C](#_Appendix_7.C:_Distribution). Data used in [appendix 7.C](#_Appendix_7.C:_Distribution) includes all tested students in the full student population. The “N/A” notation in the tables in [appendix 7.C](#_Appendix_7.C:_Distribution) indicates a score that was not possible (raw embedded PT scores of 13 and above).

For grade five, Life Sciences and Physical Sciences embedded PTs appeared to be of similar difficulty and were easier for the students than the Earth and Space Sciences embedded PTs. Earth and Space Sciences appeared to be the most difficult science domain for the grade five students. The grade five Earth and Space Sciences embedded PTs contained three of the five most difficult items on the grade five assessment, which is reflected in the lower raw scores.

For grade eight, Life Sciences appeared to be the easiest science domain for the grade eight students, with a mean raw score of 8.2. The Life Sciences domain included the three easiest items on the grade eight assessment. The Physical Sciences and Earth and Space Sciences embedded PTs appeared to be of similar difficulty for the grade eight students. Version 2 of the Physical Sciences embedded PT appeared to be the most difficult embedded PT for the grade eight students.

For the high school grades, Physical Sciences appeared to be the easiest science domain, while the Earth and Space Sciences embedded PTs appeared to be the most difficult for the high school students. This pattern was consistent across the three high school grades. The Earth and Space Sciences domain had several of the harder items on the high school assessment (four of the six hardest items on the assessment).

Table 7.3 Raw Score Summary for Each Embedded PT

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Module | Number of Students | Number of Items | Maximum Number of Points | Mean Raw Score | SD Raw Score | Minimum Raw Score | Maximum Raw Score | Mean Raw Score as a Percentage |
| **Grade 5 Total:** | 4,414 | 30 | 36 | 23.8 | 6.5 | 0 | 36 | 66.1 |
| Grade 5 PT 1 Version 1 (Life Sciences) | 2,290 | 10 | 12 | 8.3 | 2.4 | 0 | 12 | 69.2 |
| Grade 5 PT 2 Version 1 (Physical Sciences) | 2,291 | 10 | 12 | 8.6 | 2.7 | 0 | 12 | 71.7 |
| Grade 5 PT 3 Version 1 (Earth and Space Sciences) | 2,291 | 10 | 12 | 6.9 | 2.4 | 0 | 12 | 57.5 |
| Grade 5 PT 1 Version 2 (Life Sciences) | 2,124 | 10 | 12 | 8.6 | 2.5 | 0 | 12 | 71.7 |
| Grade 5 PT 2 Version 2 (Physical Sciences) | 2,123 | 10 | 12 | 8.4 | 2.7 | 0 | 12 | 70.0 |
| Grade 5 PT 3 Version 2 (Earth and Space Sciences) | 2,123 | 10 | 12 | 6.9 | 2.5 | 0 | 12 | 57.5 |
| **Grade 8 Total:** | 4,520 | 30 | 36 | 23 | 6.9 | 0 | 36 | 63.9 |
| Grade 8 PT 1 Version 1 (Life Sciences) | 2,357 | 10 | 12 | 8.2 | 2.3 | 0 | 12 | 68.3 |
| Grade 8 PT 2 Version 1 (Physical Sciences) | 2,357 | 10 | 12 | 7.5 | 2.7 | 0 | 12 | 62.5 |
| Grade 8 PT 3 Version 1 (Earth and Space Sciences) | 2,357 | 10 | 12 | 7.4 | 2.9 | 0 | 12 | 61.7 |
| Grade 8 PT 1 Version 2 (Life Sciences) | 2,163 | 10 | 12 | 8.2 | 2.5 | 0 | 12 | 68.3 |
| Grade 8 PT 2 Version 2 (Physical Sciences) | 2,163 | 10 | 12 | 7.1 | 2.5 | 0 | 12 | 59.2 |
| Grade 8 PT 3 Version 2 (Earth and Space Sciences) | 2,163 | 10 | 12 | 7.6 | 2.7 | 0 | 12 | 63.3 |
| **High School Total:** | 7,264 | 30 | 36 | 22.7 | 6.8 | 0 | 36 | 63.1 |
| HS PT 1 Version 1 (Life Sciences) | 3,836 | 10 | 12 | 7.6 | 2.6 | 0 | 12 | 63.3 |
| HS PT 2 Version 1 (Physical Sciences) | 3,836 | 10 | 12 | 7.9 | 2.6 | 0 | 12 | 65.8 |
| HS PT 3 Version 1 (Earth and Space Sciences) | 3,836 | 10 | 12 | 7.1 | 2.5 | 0 | 12 | 59.2 |
| HS PT 1 Version 2 (Life Sciences) | 3,428 | 10 | 12 | 7.9 | 2.7 | 0 | 12 | 65.8 |
| HS PT 2 Version 2 (Physical Sciences) | 3,428 | 10 | 12 | 8.1 | 2.7 | 0 | 12 | 67.5 |
| HS PT 3 Version 2 (Earth and Space Sciences) | 3,428 | 10 | 12 | 6.9 | 2.5 | 0 | 12 | 57.5 |
| **Grade 10 Total:** | 317 | 30 | 36 | 23.1 | 7.2 | 1 | 36 | 64.2 |
| Grade 10 PT 1 Version 1 (Life Sciences) | 120 | 10 | 12 | 8.1 | 2.5 | 0 | 12 | 67.5 |
| Grade 10 PT 2 Version 1 (Physical Sciences) | 120 | 10 | 12 | 8.2 | 2.5 | 0 | 12 | 68.3 |
| Grade 10 PT 3 Version 1 (Earth and Space Sciences) | 120 | 10 | 12 | 7.3 | 2.6 | 1 | 12 | 60.8 |
| Grade 10 PT 1 Version 2 (Life Sciences) | 197 | 10 | 12 | 7.8 | 2.8 | 0 | 12 | 65.0 |
| Grade 10 PT 2 Version 2 (Physical Sciences) | 197 | 10 | 12 | 8.0 | 3.0 | 0 | 12 | 66.7 |
| Grade 10 PT 3 Version 2 (Earth and Space Sciences) | 197 | 10 | 12 | 6.9 | 2.5 | 0 | 12 | 57.5 |
| **Grade 11 Total:** | 2,289 | 30 | 36 | 22.9 | 6.8 | 0 | 36 | 63.6 |
| Grade 11 PT 1 Version 1 (Life Sciences) | 1,203 | 10 | 12 | 7.8 | 2.6 | 0 | 12 | 65.0 |
| Grade 11 PT 2 Version 1 (Physical Sciences) | 1,203 | 10 | 12 | 7.9 | 2.6 | 0 | 12 | 65.8 |
| Grade 11 PT 3 Version 1 (Earth and Space Sciences) | 1,203 | 10 | 12 | 7.2 | 2.6 | 0 | 12 | 60.0 |
| Grade 11 PT 1 Version 2 (Life Sciences) | 1,086 | 10 | 12 | 7.8 | 2.7 | 0 | 12 | 65.0 |
| Grade 11 PT 2 Version 2 (Physical Sciences) | 1,086 | 10 | 12 | 8.1 | 2.7 | 0 | 12 | 67.5 |
| Grade 11 PT 3 Version 2 (Earth and Space Sciences) | 1,086 | 10 | 12 | 6.9 | 2.5 | 0 | 12 | 57.5 |
| **Grade 12 Total:** | 4,658 | 30 | 36 | 22.6 | 6.7 | 0 | 36 | 62.8 |
| Grade 12 PT 1 Version 1 (Life Sciences) | 2,513 | 10 | 12 | 7.6 | 2.6 | 0 | 12 | 63.3 |
| Grade 12 PT 2 Version 1 (Physical Sciences) | 2,513 | 10 | 12 | 7.8 | 2.5 | 0 | 12 | 65.0 |
| Grade 12 PT 3 Version 1 (Earth and Space Sciences) | 2,513 | 10 | 12 | 7.0 | 2.5 | 0 | 12 | 58.3 |
| Grade 12 PT 1 Version 2 (Life Sciences) | 2,145 | 10 | 12 | 7.9 | 2.7 | 0 | 12 | 65.8 |
| Grade 12 PT 2 Version 2 (Physical Sciences) | 2,145 | 10 | 12 | 8.1 | 2.7 | 0 | 12 | 67.5 |
| Grade 12 PT 3 Version 2 (Earth and Space Sciences) | 2,145 | 10 | 12 | 6.9 | 2.5 | 0 | 12 | 57.5 |

### DIF Analyses

DIF analyses were conducted for 2018–2019 CAA for Science items with sufficient sample sizes. The minimum sample size requirements for the DIF analyses were 400 in the combined focal and reference groups and 100 in the smaller of the two groups. These sample sizes are based on standard operating procedures for DIF analyses at ETS.

If an item performs differentially across identifiable student groups (e.g., gender or ethnicity) when students are matched on ability, the item may be measuring something other than the intended construct. 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.

#### Dichotomous Items

The Mantel-Haenszel (MH) DIF statistic was calculated for dichotomous items (Mantel & Haenszel, 1959; Holland & Thayer, 1985). Using the total raw score as the criterion score, students at each raw score level in the focal group (e.g., Hispanic students) are compared with examinees at the same raw score level in the reference group (e.g., non-Hispanic White students). The common odds ratio is estimated across the total raw score using the formula in equation 7.4 (Dorans & Holland, 1993). The resulting estimate is interpreted as the relative likelihood of success on a particular item for members of two groups when matched on ability.

*Refer to the* [*Alternative Text for Equation 7.4*](#_Alternative_Text_for_3) *for a description of this equation.*

Equation 7.4; a link to the long description for this equation is found in the preceding paragraph. (7.4)

where,

*m* indexes the score categories,

*Rrm* is the number of students in the reference group who answer the item correctly,

*Wrm* is the number of students in the reference group who answer the item incorrectly,

*Ntm* is the total number of students,

*Rfm* is the number of students in the focal group who answer the item correctly, and

*Wfm* is the number of students in the focal group who answer the item incorrectly.

To facilitate the interpretation of MH results, the common odds ratio is transformed to the delta scale using the formula presented in equation 7.5 (Holland & Thayer, 1988). *Refer to the* [*Alternative Text for Equation 7.5*](#_Alternative_Text_for_4) *for a description of this equation.*

 (7.5)

Positive values indicate DIF in favor of the focal group (i.e., positive DIF items are differentially easier for the focal group) whereas negative values indicate DIF in favor of the reference group (i.e., negative DIF items are differentially easier for the reference group).

#### Polytomous Items

The standardization DIF (Dorans & Schmitt, 1993; Zwick, Thayer, & Mazzeo, 1997; Dorans, 2013) is used in conjunction with the Mantel chi-square statistic (Mantel, 1963; Mantel & Haenszel, 1959) to identify polytomous items with DIF; the former measures the size of the DIF while the latter indicates the significance level of the DIF. The standardized mean difference (SMD) compares the item means of the two groups after adjusting for differences in the distribution of students across the values of the matching variable. SMD is calculated using the formula presented in equation 7.6. *Refer to the* [*Alternative Text for Equation 7.6*](#_Alternative_Text_for_5) *for a description of this equation.*

 (7.6)

where,

*X* isthe criterion score,

*Y* isthe item score,

*M* is the number of score categories on *X*,

*Nrm* is the number of students in the reference group in score category *m*,

*Nfm* is the number of students in the focal group in score category *m*,

*Er* is the expected item score in the reference group, and

*Ef* is the expected item score in the focal group.

A positive SMDvalue means that after statistically matching on the criterion score, the focal group has a higher mean item score than the reference group. A negative SMD value means that after statistically matching on the criterion score, the focal group has a lower mean item score than the reference group.

#### DIF Categories and Definitions

Based on the DIF statistics and significance tests, items are classified into three categories, labeled A, B, and C. 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 categorization rules for dichotomous items are presented in table 7.4; the categorization rules for polytomous items are presented in table 7.5.

Table 7.4 DIF Categories for Dichotomous Items

|  |  |
| --- | --- |
| DIF Category | Criteria |
| A (negligible) | * Absolute value of MH D-DIF is not significantly different from zero or is less than one. * Positive values are classified as “A+” and negative values as “A-.” |
| B (moderate) | * Absolute value of MH D-DIF is significantly different from zero but not from one and is at least one; OR * Absolute value of MH D-DIF is significantly different from one but is less than 1.5. * Positive values are classified as “B+” and negative values as “B-.” |
| C (large) | * Absolute value of MH D-DIF is significantly different from one and is at least 1.5. * Positive values are classified as “C+” and negative values as “C-.” |

In table 7.5, SMD is standardized mean difference, and SD is total group standard deviation of the item score.

Table 7.5 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 |

DIF analyses were conducted on each item for designated comparison groups, if the number of students in the group was sufficient. Groups were defined on the basis of demographic variables, including gender, race or ethnicity, and primary disabilities. These comparison groups are specified in table 7.6. An asterisk (\*) indicates where DIF analysis was not performed due to insufficient sample sizes for all three CAA for Science assessments.

Table 7.6 Student Groups for DIF Comparison

|  |  |  |
| --- | --- | --- |
| DIF Type | Reference Group | Focal Group |
| **Gender** | Male | * Female |
| **Race/Ethnicity** | White | * American Indian or Alaska Native\* * Asian * Black or African American * Filipino * Hispanic or Latino * Native Hawaiian or Other Pacific Islander\* |
| **Disability** | Intellectual Disability | * Autism * Deaf-blindness\* * Emotional disturbance\* * Hearing impairment\* * Multiple disabilities * Orthopedic impairment * Other health impairment * Specific learning disability * Speech or language impairment * Traumatic brain injury\* * Visual impairment\* |
| **High School Grade Level** | Grade Eleven | * Grade ten * Grade twelve |

#### DIF Analysis Results Summary

The DIF results tables in [appendix 7.D](#_Appendix_7.D:_DIF) show the results of the DIF analyses performed on each test. These tables, which are based on the data file available in June 2019, show the number of items classified into each category for each DIF analysis performed. Items are classified as C+, B+, A+, A-, B-, C-, or N/A. N/A indicates that the DIF analysis was not performed due to insufficient sample size. The DIF analyses were not performed for several comparisons due to the small samples, which are identified in table 7.6.

In the DIF results tables, data in the *N* column shows the number of items in each classification. Out of the 124 items across the three CAA for Science assessments, only one item (or less than 1% of the items) was flagged for possible DIF.

Table 7.D.4 lists the grade five item flagged during the DIF analyses between the Intellectual Disability group and the Specific Learning Disability group (MH D-DIF value of 2.47). This item was reviewed by the data review panelists to identify possible reasons that might explain the differences between the Intellectual Disability group and the Specific Learning Disability group. The data review panelists could not identify a reason for the differences in student performance across the two groups.

### Item Response Theory (IRT) Analyses

IRT is built upon the item response function, which describes the probability of a given response as a function of a person’s true ability. IRT can be used to implement item calibrations, link item parameters, scale test scores across different forms or test administrations, evaluate item performance, build an item bank, and assemble test forms.

This section describes how IRT models are used in CAA tests for calibrating items. IRT data file preparation and IRT models are also covered in this section.

#### IRT Models

The one-parameter logistic (1PL) IRT model (1PL-IRT) is used for the CAA for Science item calibration and was selected after consultation with the CDE. The generalized partial credit model (GPCM) (Muraki, 1992) restricted for 1PL-IRT, which is essentially the partial credit model (PCM) (Masters, 1982), is applied to both dichotomous and polytomous items. The mathematical form of the GPCM is presented in equation 7.7. *Refer to the* [*Alternative Text for Equation 7.7*](#_Alternative_Text_for_6) *for a description of this equation.*

 (7.7)

where,

 is the probability of student with proficiency  obtaining score *h* on item *i*,

 is the maximum number of score points for item *i*,

 is the discrimination parameter and is fixed to 0.588 for every item,

 is the location (difficulty) parameter for item *i*,

 is the category difficulty parameter for item *i* on score *v*, and

*D* is a scaling constant of 1.7 that makes the logistic model approximate the normal ogive model.

When , equation 7.7 becomes an expression of the 1PL model for dichotomous items. Essentially, the 1PL model (Hambleton, Swaminathan, & Rogers, 1991) and the PCM (Masters, 1982) were used for dichotomous items and polytomous items, respectively.

#### Item Calibration

After the 2018–2019 CAA for Science administration, all the items within each grade-level assessment were calibrated concurrently, using all available data. Previous studies show that compared with separate calibration, concurrent calibration is more accurate when the data fits the IRT model (Kim & Cohen, 1998; Hanson & Béguin, 2002). After consultation with the CDE, a single-group concurrent calibration approach was used for item calibration of the CAA for Science. As stated in subsection [*7.6.1 IRT Models*](#_IRT_Models)*,* the 1PL model (Hambleton, Swaminathan, & Rogers, 1991) and the corresponding PCM were jointly used to concurrently calibrate dichotomously and polytomously scored items*.* The software flexMIRT® (Cai, 2016) version 3.5 was used for calibration.

##### 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 were reviewed also by ETS content experts and the CDE. The decision to remove items from calibration were made in consultation with the CDE. For the 2018–2019 administration of the CAA for Science, no items were excluded from the calibration analyses.

For IRT calibration, scored item response data was used to create the IRT analysis input data files for each grade and content area, including responses to items for both versions of the CAA for Science. The IRT analysis input data file is a sparse matrix, because each student completed only one of the two possible versions. Similar to the classical item analyses, “omit” items are treated as incorrect and “not-presented” items are treated as blank.

##### Description of the Calibration Procedure

FlexMIRT (Cai, 2016), a multilevel and multiple-group IRT software package for item analysis and test scoring, is used for CAA for Science 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.

The calibration procedure is as follows:

1. Receive test form planners and create the item mapping files
2. Receive data
3. Run complete classical item analysis and create the sparse matrices
4. Create the flexMIRT control files
5. Run flexMIRT and evaluate the results

The procedure described next was followed to calibrate the 2018–2019 student response data using flexMIRT for each grade and subject:

1. Prepare and format the input data files as required by flexMIRT
2. Prepare flexMIRT control files
3. Evaluate the flexMIRT output to examine whether every execution of flexMIRT analysis reached satisfactory convergence
4. Review the item parameter estimates
   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 model-fit. The model-fit was evaluated using the root mean square error of approximation (RMSEA). RMSEA values less than 0.05 indicate good fit while RMSEA values greater than 0.10 indicate poor fit (Browne & Cudeck, 1993). The *b*-parameters were correlated with the *p*-values.
   2. At the item level, statistics of individual items were examined, including item difficulty estimates (*b*-parameters and *d*-parameters), item-fit statistics using the marginal chi-square statistic. The *b*-parameters and the *d*-parameters should be in the range of -4.0 to +4.0 with a standard error of 0.10 or less.
5. Flag items that did not perform as expected
   1. 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 conducted independently by two ETS psychometricians to ensure quality and accuracy of results. The 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 subsection [*10.4 Quality Control of Psychometric Processes*](#_Quality_Control_of)for more details of this procedure.

##### Summary of the Calibration Results

The overall summary of IRT *b*-value estimates for the CAA for Science field test calibration analyses is shown in table 7.E.1. The mean SD, minimum, and maximum values are presented, in addition to the number of items for each test. The RMSEA values are also provided in table 7.E.1, which were 0.03 for grades five and eight and 0.04 for high school. These values are all below the value 0.05 which indicates good model fit (Browne & Cudeck, 1993). All the item parameters were between -4.0 and +4.0. The average *b*‑parameter for all three CAA for Science assessments were below zero, indicating that, in general, the items were relatively easy for these students.

Table 7.E.2 through table 7.E.4 provide the IRT difficulty and step parameter estimates at the item level for each CAA for Science test. Table 7.E.5 through table 7.E.7 provide the *p*‑value, *b*-value, and omit rate for each item.

As a result of consultation with the CDE, no items used during the 2018–2019 CAA administration were removed from the analysis and no categories were collapsed.

### Reliability Analyses

Reliability is the extent to which differences in test scores reflect true differences in the knowledge, ability, or skill being tested rather than fluctuations due to measurement error. Thus, reliability is the consistency of scores across conditions that do not differ systematically and only contain random measurement errors. In statistical terms, the variance in the distributions of test scores—essentially, the differences among individuals—is due partly to real differences in the knowledge, skill, or ability being tested (true variance) and due partly to measurement error inherent in the measurement process (error variance). The reliability coefficient is an estimate of the proportion of the total variance that is true variance.

Reliability coefficients can possibly range from 0 to 1. The higher the reliability coefficient for a set of scores, the more likely that the students would obtain very similar scores upon repeated testing occasions (assuming there is no memory or practice effect) if the students did not change in their level of the knowledge or skills measured by the test.

There are several different ways of estimating reliability. Stratified Alpha was computed for the reliability estimates for each version of the CAA for Science for the student groups. More details can be found in the next subsection, [7.7.1](#_Internal_Consistency_Reliability).

The standard error of measurement (SEM) is a measure of the extent to which students’ scores tend to differ from their true scores. A student’s true score can be thought of as the mean observed scores a student would earn over an infinite number of independent administrations of the test. The larger the SEM, the more the variability of a student’s observed scores across repeated testing. Observed scores with large SEMs pose a challenge to the valid interpretation of test scores.

#### Internal Consistency Reliability

In classical test theory, the reliability coefficient can be defined as the squared correlation between the observed score and the true score, which is equal to the correlation between parallel observed scores (Lord and Novick, 1968, p.61). In applied settings, the requirement of repeated administrations is impractical, and methodologies estimating reliability from relationships among student performances on items within a single test form are often used. Coefficient alpha (Cronbach, 1951) is among the most common of these methodologies.

The formula for the internal consistency reliability as measured by Cronbach’s Alpha (Cronbach, 1951) is presented in equation 7.8. *Refer to the* [*Alternative Text for Equation 7.8*](#_Alternative_Text_for_7) *for a description of this equation.*

Equation 7.8; a link to the long description for this equation is found in the preceding paragraph. (7.8)

where,

*n* is the number of items,

sigma sub i, squared is the variance of scores on the *i*-th item, and

sigma sub x, squared is the variance of the total score (sum of scores on the individual items).

Since the CAA for Science assessments have a mix of item types (both dichotomous and polytomous items), it is more appropriate to report stratified Alpha (Feldt & Brennan, 1989). The stratified Alpha is a weighted average of Cronbach’s Alpha for item sets with different maximum score points or “strata.” The item sets used when calculating the stratified Alpha are dichotomous and polytomous items for each of the three tests.

The formula for calculating the stratified Alpha is presented in equation 7.9. *Refer to the* [*Alternative Text for Equation 7.9*](#_Alternative_Text_for_8) *for a description of this equation.*

Equation 7.9; a link to the long description for this equation is found in the preceding paragraph. (7.9)

where,

 is the variance for stratum *j* of the test,

 is the total variance of the test, and

 is the Cronbach’s Alpha for stratum *j* of the test.

#### Standard Error of Measurement

The SEM provides a measure of score instability on the raw score metric. The SEM is the square root of the error variance in the scores (i.e., the SD of the distribution of the differences between students’ observed scores and their true scores). The SEM is calculated using the formula presented in equation 7.10. *Refer to the* [*Alternative Text for Equation 7.10*](#_Alternative_Text_for_9) *for a description of this equation.*

 (7.10)

where,

 is the reliability estimated in equation 7.9, and

 is the SD of the total score (raw score).

The SEM is useful in determining the confidence interval that likely captures a student’s true score. A student’s true score can be thought of as the mean of observed scores a student would earn over an infinite number of independent administrations of the assessment. Approximately 95 percent of the students will have scores within the range of their true scores: -1.96 SEMs to their true scores +1.96 SEMs (Crocker & Algina, 1986). For example, if a student’s observed score on a given test equals 345 points, and the SEM equals five, one can be 95 percent confident that the student’s true score lies between 335 and 355 points (i.e., 345 ± 10).

Table 7.7 presents the total score reliability for the raw scores, and the mean, SD, and SEM for each test, along with the number of students upon whose responses those analyses were performed. Note that the reliability is for raw scores on the whole test. The reliabilities range from 0.84 to 0.87.

Table 7.7 Summary Statistics for Raw Scores and Stratified Alpha

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level and Version | No. of Operational Items | No. of Students | Stratified Alpha | Raw Score SEM | Mean Raw Score | SD Raw Score |
| Grade 5 Version 1 | 30 | 2,196 | 0.84 | 2.52 | 23.84 | 6.39 |
| Grade 5 Version 2 | 30 | 2,035 | 0.86 | 2.43 | 23.86 | 6.70 |
| Grade 8 Version 1 | 30 | 2,255 | 0.87 | 2.47 | 23.15 | 6.96 |
| Grade 8 Version 2 | 30 | 2,063 | 0.86 | 2.46 | 22.82 | 6.72 |
| High school Version 1 | 30 | 3,694 | 0.85 | 2.53 | 22.60 | 6.63 |
| High school Version 2 | 30 | 3,330 | 0.86 | 2.53 | 22.89 | 6.88 |
| Grade 10 Version 1 | 30 | 110 | 0.85 | 2.50 | 23.62 | 6.57 |
| Grade 10 Version 2 | 30 | 189 | 0.88 | 2.50 | 22.71 | 7.48 |
| Grade 11 Version 1 | 30 | 1,148 | 0.85 | 2.50 | 22.94 | 6.79 |
| Grade 11 Version 2 | 30 | 1,058 | 0.86 | 2.53 | 22.80 | 6.86 |
| Grade 12 Version 1 | 30 | 2,436 | 0.84 | 2.54 | 22.38 | 6.55 |
| Grade 12 Version 2 | 30 | 2,083 | 0.86 | 2.53 | 22.96 | 6.84 |

All of the test versions have reliability greater than 0.80, a value that is considered acceptable for most assessments. The reliability estimates and SEMs are provided in table 7.F.1 through table 7.F.6 for the student groups based on gender, ethnicity, English language fluency, economic status, migrant status, primary disability, accommodation, and designated support.

Most student groups have reliability greater than 0.80 across all assessments with the exceptions of Native Hawaiian or Other Pacific Islander, reclassified fluent English proficient, and some disability groups for some grades and some test versions. The lowest reliability is speech or language impairment for grade eleven version 1, with the reliability coefficient of 0.72. It should be noted that in this case, the low reliability was likely due to the lack of variation in student performance in relation to the homogeneous groups and small group size.

### 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 tests and the uses and interpretations of test scores.

The validation process does not rely on a single study or gathering only one type of evidence. Rather, validation involves multiple investigations and different kinds of supporting evidence (AERA, APA, & NCME, 2014; Cronbach, 1971; ETS, 2014; Kane, 2006). It begins with the test design and is implicit throughout the entire assessment process, which includes item development and field testing, analyses of items, test scaling and 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 CAA for Science. This section is organized primarily around the principles prescribed by AERA, APA, and NCME’s *Standards for Educational and Psychological Testing* (2014). These *Standards* require a clear definition of the purpose of the test, a description of the constructs to be assessed, and the population to be assessed, as well as how the scores are to be interpreted and used.

The *Standards* identify five kinds of evidence that can provide support for score interpretations and uses:

1. Evidence based on test content
2. Evidence based on 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 CAA for Science, followed by a description and discussion of different kinds of validity evidence that have been gathered.

#### Evidence in the Design of the CAA for Science

##### Purpose

The CAA for Science is designed to assess students with the most significant cognitive disabilities whose individualized education program (IEP) teams have designated the use of an alternate assessment on the statewide summative assessments.

The CAA for Science is designed to show how well students perform relative to the California Next Generation Science Standards (CA NGSS) Core Content Connectors (Science Connectors), which were developed by the National Center and State Collaborative. These Science Connectors are content targets linked to the CA NGSS and yet are less complex than the CA NGSS, while focusing on the main academic content at each subject and grade level.

##### The Constructs to Be Measured

The Science Connectors illustrate the necessary knowledge and skills needed to reach the learning targets within the CA NGSS and the knowledge and skills needed at each grade level. The Science Connectors identify priorities in each content area to guide instruction for students in this population and for the alternate assessment.

Test blueprints are used to measure the Science Connectors (CDE, 2018a). They also provide an operational definition of the construct to which each set of standards refers and define the following for each content area:

* 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) in order to minimize construct-irrelevant score variance (Messick, 1989).

ETS developed all CAA for Science test items to conform to the SBE-approved Science Connectors and test blueprints.

##### Intended Test Population

Only eligible students may participate in the administration of the CAA for Science. Any student identified for alternate testing in grades five and eight and high school (grade ten, eleven, or twelve) takes the CAA for Science. IEP teams “shall determine when a child with a significant cognitive disability shall participate in an alternate assessment aligned with the alternate academic achievement standards.”[[6]](#footnote-7)

#### Evidence Based on Test Content

Evidence based on test content refers to traditional forms of content validity evidence, such as the rating of test specifications and test items (Crocker, Miller, & Franks, 1989; Sireci, 1998), as well as alignment methods for educational tests that evaluate the interactions between curriculum frameworks, testing, and instruction (Rothman, Slattery, Vranek, & Resnick, 2002; Bhola, Impara, & Buckendahl, 2003; Martone & Sireci, 2009).

[*Chapter 4: Test Assembly*](#_Test_Assembly) contains information on which the test forms administered in 2018–2019 were built.

##### Description of the State Standards

The CAA for Science is aligned with the NGSS Science Connectors. The Science Connectors illustrate the necessary knowledge and skills needed to reach the learning targets within the CA NGSS and the knowledge and skills needed in each grade. They also identify priorities in each content area to guide the instruction for students in this population and for the alternate assessment (2018b).

##### Embedded PT and Item Specifications

Item specifications describe the characteristics of items that are written to measure each content standard. The specifications for science are described in [*Chapter 3: Embedded PT Development and Review*](#_Embedded_PT_Development).

##### Assessment Blueprints

The CAA for Science test blueprints describe the content of the CAA for Science for all grades tested and how that content is assessed (CDE, 2018a). The test blueprints address the basic core content domains, the CA NGSS, the Science Connectors, and the essential understanding for each standard. Each test is described by a single blueprint. A description of the test blueprint is provided in [*Chapter 4: Test Assembly*](#_Test_Assembly).

##### Form Assembly Process

The content standards and blueprints are the basis for choosing items for each assessment. Additionally, item difficulty and the content complexity of items are provided to evaluate the statistical characteristics of test forms. Refer to [*Chapter 4: Test Assembly*](#_Test_Assembly) for information on the test assembly process.

#### Evidence Based on Response Processes

Validity evidence based on response processes refers to “evidence concerning the fit between the construct and the detailed nature of performance or response actually engaged in by students” (AERA et al., 2014, p. 12). This type of evidence generally includes documentation of activities such as

* systematic observations of test response behavior;
* showing the relationships of items intended to require demonstrations or applications of knowledge and skills to other measures that require similar levels of cognitive complexity in the content (i.e., teacher ratings of student performance); 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 CAA for Science is measuring the cognitive skills that are intended as the objects of measurement and are used by students to respond to the items. For example, the survey questions administered after each embedded PT were analyzed as part of the research agenda with the goal of understanding the CAA for Science embedded PTs. A summary of the student survey results are provided in subsection [*8.1.1 Student Survey*](#_Student_Survey).

##### Analysis of Testing Time

Testing time for each 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 a test is collected and analyzed to build a profile describing what a typical testing event looks like for each content area and grade. 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 CAA for Science are untimed tests.

The students with no item response and students who did not answer at least one item from each of the three embedded PTs were removed from these analyses. The remaining testing population is partitioned into quartiles based on raw scores. These quartile groupings are not the same as the achievement levels.

Descriptive statistics of the time required to complete the total test are computed for each of the four quartile groups for each assessment. Because some cases of extremely long testing time may be attributed to taking longer to complete the assessments or that the assessment was not closed down properly, the results should be interpreted with caution. The medians (50th percentile) are more meaningful in the interpretation of the time comparisons because medians are less impacted by extreme values than means.

Table 7.G.8 provides summary statistics and selected percentiles of the distribution of total testing time for each version of the test at each grade level. Table 7.G.9 through table 7.G.14 provide this information at each raw score quartile level. The unit of testing time is minutes; for example, in table 7.G.9, the median of the testing time for the first quartile group (Q1) for grade five version 1 is 5.25 minutes.

The median testing time ranges from 9.93 minutes for grade five version 1 to 12.86 minutes for grade twelve version 2. Overall, students at the lowest quartile level (Q1) had shorter testing times than students in the other quartile groups. For grades five and eight, students in quartile 4 (Q4) had the longest testing times. However, at the high school grades, students in quartile 3 (Q3) had the longest testing times.

For grade five version 1, students in quartile 2 (Q2) and Q3 groups had similar median testing times of 10.41 and 10.40, respectively. At grade five version 2, the median total testing time was similar for the Q2, Q3, and Q4 groups, ranging from 11.17 minutes to 11.75 minutes. At grade eight, the median testing time increased from Q1 through to Q4. For grades five and eight, students who spent more time completing the three embedded PTs have higher raw scores.

The relationship between testing time and student performance at the high school level is not as clear as it is for grades five and eight. At high school, the median total testing time and student performance increased from Q1 to Q3. Although students’ raw scores increased for Q4, the time high school students spent completing the three embedded PTs decreased slightly.

Overall, the testing time students used to complete the three embedded PTs is lower than the testing time required by students taking the CAAs for ELA and mathematics (CDE, 2020, appendix 8.G).

#### Evidence Based on Internal Structure

Internal structure evidence evaluates the strength or salience of the major dimensions underlying an assessment using indices of measurement precision such as DIF analysis, test reliability, and SEMs.

##### DIF

DIF analyses were conducted to assess differences in the item performance of groups of students who differ in their demographic characteristics. For the CAA for Science, only one item was identified as having significant levels of DIF. Refer to subsection [*7.5 DIF Analyses*](#_DIF_Analyses) for a description the DIF analyses and [appendix 7.D](#_Appendix_7.D:_DIF), where the results of the DIF analyses are reported.

##### Overall Reliability Estimates

The results of reliability analyses on the raw scores for each test are presented in subsection [*7.7 Reliability Analyses*](#_Reliability_Analyses). The results indicate that the stratified alpha reliability estimates for all tests are moderately high, ranging from 0.84 to 0.87.

##### Student Group Reliability Estimates

The reliabilities are also examined for various student groups. The student groups considered are based on gender, ethnicity, economic status, primary disability, migrant status, and English language fluency. Reliability estimates and SEM information for the raw scores are reported for each student group in table 7.F.1 through table 7.F.6 in [appendix 7.F](#_Appendix_7.F:_Reliability). Reliability estimates and SEM information for the raw scores are reported for each student group in table 7.F.1 through table 7.F.6 in [appendix 7.F](#_Appendix_7.F:_Reliability).

#### Evidence Based on Relationship to Other Variables

Evidence based on *relations to other variables* can be evaluated using the correlation between the CAA for Science assessment results and variables related to students. Two variables analyzed related to the students’ results include the CAAs for ELA and mathematics and the level of student engagement while taking the three embedded PTs.

The correlations between the CAA for Science and the CAAs for ELA and mathematics range from 0.79 to 0.86. Table 7.8 provides the correlations of the raw scores on the CAA for Science with the scale scores on the CAAs for ELA and mathematics. The number of students on which these correlations are based are provided in table 7.8 and are based on students having a valid test score for all three CAAs. At the high school level, the correlations are computed only for grade eleven students who completed all three CAAs during the 2018–2019 CAASPP administration.

In general, the correlations between the CAA for Science and the CAAs for ELA, as well as between the CAA for Science and mathematics, were high.

Table 7.8 Correlations Among the CAAs for ELA, Mathematics, and Science

|  |  |  |  |
| --- | --- | --- | --- |
| Grade Level | Number of Students | Correlation between Science and ELA | Correlation Between Science and Math |
| Grade Five | 5,079 | 0.85 | 0.80 |
| Grade Eight | 5,179 | 0.81 | 0.79 |
| Grade Eleven | 2,609 | 0.86 | 0.83 |

Table 7.G.1 presents the relationship between scale scores and the level of test engagement reported by teachers on a Likert scale of 0–3. Refer to subsection[*8.1.1 Student Survey*](#_Student_Survey) for additional information.

Results show a moderate correlation between the test engagement and student achievement (raw scores). As is shown by their test engagement, high school students seem to have a higher level of test engagement, on average, than students in grades five and eight.

Table 7.G.2 through table 7.G.7 present the mean raw score for each embedded PT by level of student engagement. For all embedded PTs, students who were fully engaged had the highest mean raw scores compared to students who were either moderately engaged or minimally engaged. Students who were minimally engaged had the lowest mean raw score across all embedded PTs and grade levels.

The minimal levels of engagement for some students could be related to whether students could access the test content while testing or whether the student had the opportunity to learn the content prior to testing. The student’s familiarity with the content or tasks would also impact the student’s level of engagement. Some students might be more engaged with familiar content, while other students might be more engaged when the content or task is unique (i.e., a novel experience).

### 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. (2016). FlexMIRT®: *Flexible multilevel, multidimensional item analysis and test scoring* (Version 3.5) [computer software]. Chapel Hill, NC: Vector Psychometric Group.

California Department of Education. (2018a). *California Alternate Assessment for Science blueprint*. Sacramento, CA: California Department of Education.

California Department of Education. (2018b). *California Next Generation Science Standards Core Content Connectors for Alternate Assessments.* Sacramento, CA: California Department of Education.

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

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*(3), 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). Phoenix, AZ: Oryx Press.

Green, B. F., Bock, R. D., Humphreys, L. G., Linn, R. L., & Reckase, M. D. (1984). Technical guidelines for assessing computerized adaptive tests. *Journal of Educational Measurement*, 21(4), 347–360.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications.

Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common item equating design. *Applied Psychological Measurement, 26*, 3–24.

Holland, P., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.

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.

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

Kearns, J., Kleinert, H., Kleinert, J., and Towles-Reeves, E. (2006). Learner characteristics inventory. Lexington, Kentucky: University of Kentucky, National Alternate Assessment Center.

Kim, S. H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement, 22*, 131–143.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley Publishing Company.

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), 149–‍174.

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*(2): 159–76.

Nunnally, J. C. (1978). *Psychometric Theory* (2nd ed.). New York: McGraw-Hill.

Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (2002). *Benchmarking and alignment of standards and testing* [Technical Report 566]. Washington, DC: Center for the Study of Evaluation.

Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment, 5*, 299–321.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education, 10*(4), 321–44.

### Accessibility Information

#### Alternative Text for Equation 7.1

P value sub dich equals the fraction with the numerator the sum of X sub ij and the denominator N sub I end fraction. *(Return to* [*equation 7.1*](#equation_7_1)*.)*

#### Alternative Text for Equation 7.2

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. *(Return to* [*equation 7.2*](#equation_7_2)*.)*

#### Alternative Text for Equation 7.3

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. *(Return to* [*equation 7.3*](#equation_7_3)*.)*

#### Alternative Text for Equation 7.4

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. *(Return to* [*equation 7.4*](#equation_7_4)*.)*

#### Alternative Text for Equation 7.5

MH D-DIF equals negative 2.35 times the natural logarithm open bracket alpha sub MH close bracket. *(Return to* [*equation 7.5*](#equation_7_5)*.)*

#### Alternative Text for Equation 7.6

SMD equals the fraction with numerator the sum from m equals 1 to M of N sub fm times E sub f of Y from X equals m and denominator the sum from m equals 1 to M of N sub fm end fraction minus the fraction with numerator the sum from m equals 1 to M of N sub fm times E sub r of Y from X equals m and denominator the sum from m equals 1 to M of N sub fm end fraction equals the fraction with the numerator the sum from m equals 1 to M of D sub fm and the denominator m equals1 to M of N suf fm end fraction. *(Return to* [*equation 7.6*](#equation_7_6)*.)*

#### Alternative Text for Equation 7.7

P sub ih of theta sub j equals:

The numerator exp open parenthesis the sum from v equals 1 to h of 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. *(Return to* [*equation 7.7*](#equation_7_7)*.)*

#### Alternative Text for Equation 7.8

alpha equals the fraction n over n minus one, times one minus the fraction the sum from i equals one to n, of sigma sub i, squared, over sigma sub x, squared. *(Return to* [*equation 7.8*](#equation_7_8)*.)*

#### Alternative Text for Equation 7.9

stratified alpha equals one minus the fraction of the sum sigma sub xj, squared times one minus the alpha sub j over sigma sub x, squared. *(Return to* [*equation 7.9*](#equation_7_9)*.)*

#### Alternative Text for Equation 7.10

SEM equals s sub t times the square root of 1 minus rho of theta hat theta hat prime. *(Return to* [*equation 7.10*](#equation_7_10)*.)*

### Appendix 7.A: Classical Item Analyses

**Note 1:** Items with poor statistics are flagged. Refer to the table, next, for a description of each flag and possible values that will appear in the *Flag* column in table 7.A.2 through table 7.A.7.

|  |  |  |
| --- | --- | --- |
| Flag | Description | Criteria |
| A | Indicates low average item score (AIS) /low *p*-value (difficult item) | Dichotomous item: *p*-value < 0.33  Polytomous item: AIS < 30 percent of maximum possible score points |
| H | Indicates high average item score (AIS) /high *p*‑value (easy item) | Dichotomous item: *p*-value > 0.95  Polytomous item: AIS > 80 percent of maximum possible score points |
| Rpoly | Indicates low correlation with the criterion | Item – Total Correlation < 0.20 |
| O | Indicates high percent of omits or not responding | Dichotomous item: %omit > 5%  Polytomous item: %omit > 15% |
| D | Indicates high ability students selected a distractor | Dichotomous item: High scoring students tend to select a distractor more often than the correct option  Polytomous item: High scoring students tend to score lower than at the top score level (0 or 1 for 2-point item) |
| L | Indicates a small percentage of students obtaining a score category | Polytomous item: percentage obtaining any score category lower than or equal to 3% |

**Note 2:** Items that do not have a flagged are indicated with “[no flag]” in the *Flag* column in table 7.A.2 through table 7.A.7.

Table 7.A.1 Classical Item Statistics for Each Embedded PT

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level and Version | Number of Unique Items | Maximum Number of Students | Mean *p-*value | Minimum *p-*value | Maximum *p-*value | Mean Item-Total Polyserial Correlation |
| **Grade 5 Version 1 Total:** | 30 | 2,290 | 0.67 | 0.38 | 0.92 | 0.55 |
| Grade 5 PT 1 (Life Sciences) Version 1 | 10 | 2,290 | 0.73 | 0.39 | 0.92 | 0.58 |
| Grade 5 PT 2 (Physical Sciences) Version 1 | 10 | 2,290 | 0.72 | 0.41 | 0.88 | 0.60 |
| Grade 5 PT 3 (Earth and Space Sciences) Version 1 | 10 | 2,290 | 0.58 | 0.38 | 0.84 | 0.45 |
| **Grade 5 Version 2 Total:** | 30 | 2,123 | 0.67 | 0.35 | 0.91 | 0.58 |
| Grade 5 PT 1 (Life Sciences) Version 2 | 10 | 2,123 | 0.75 | 0.49 | 0.91 | 0.64 |
| Grade 5 PT 2 (Physical Sciences) Version 2 | 10 | 2,123 | 0.70 | 0.39 | 0.86 | 0.62 |
| Grade 5 PT 3 (Earth and Space Sciences) Version 2 | 10 | 2,123 | 0.58 | 0.35 | 0.83 | 0.48 |
| **Grade 8 Version 1 Total:** | 30 | 2,357 | 0.66 | 0.39 | 0.93 | 0.58 |
| Grade 8 PT 1 (Life Sciences) Version 1 | 10 | 2,357 | 0.70 | 0.46 | 0.93 | 0.57 |
| Grade 8 PT 2 (Physical Sciences) Version 1 | 10 | 2,357 | 0.65 | 0.43 | 0.81 | 0.58 |
| Grade 8 PT 3 (Earth and Space Sciences) Version 1 | 10 | 2,357 | 0.63 | 0.39 | 0.87 | 0.60 |
| **Grade 8 Version 2 Total:** | 30 | 2,163 | 0.64 | 0.29 | 0.93 | 0.56 |
| Grade 8 PT 1 (Life Sciences) Version 2 | 10 | 2,163 | 0.69 | 0.37 | 0.93 | 0.58 |
| Grade 8 PT 2 (Physical Sciences) Version 2 | 10 | 2,163 | 0.60 | 0.29 | 0.78 | 0.53 |
| Grade 8 PT 3 (Earth and Space Sciences) Version 2 | 10 | 2,163 | 0.63 | 0.37 | 0.85 | 0.57 |
| **High School Version 1 Total:** | 30 | 3,836 | 0.65 | 0.33 | 0.90 | 0.55 |
| HS PT 1 (Life Sciences) Version 1 | 10 | 3,836 | 0.66 | 0.33 | 0.90 | 0.59 |
| HS PT 2 (Physical Sciences) Version 1 | 10 | 3,836 | 0.67 | 0.50 | 0.81 | 0.55 |
| HS PT 3 (Earth and Space Sciences) Version 1 | 10 | 3,836 | 0.60 | 0.44 | 0.79 | 0.50 |
| **High School Version 2 Total:** | 30 | 3,428 | 0.64 | 0.33 | 0.90 | 0.56 |
| HS PT 1 (Life Sciences) Version 2 | 10 | 3,428 | 0.67 | 0.33 | 0.90 | 0.60 |
| HS PT 2 (Physical Sciences) Version 2 | 10 | 3,428 | 0.69 | 0.59 | 0.80 | 0.60 |
| HS PT 3 (Earth and Space Sciences) Version 2 | 10 | 3,428 | 0.57 | 0.36 | 0.75 | 0.48 |

Table 7.A.2 Average Item Score and Polyserial Correlation for Each Item, Grade Five

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Item ID | Item Use | Version | Flag | Sample Size | AIS | Polyserial | Omit Rate | Maximum Score Points | Item Type |
| VH861679 | FT | 1 | Rpoly | 2,279 | 0.40 | 0.12 | 1% | 1 | MCSS–Discrete |
| VH861685 | FT | 2 | [no flag] | 2,108 | 0.78 | 0.64 | 2% | 1 | MCSS–Discrete |
| VH861690 | FT | Both | [no flag] | 4,388 | 0.67 | 0.41 | 2% | 1 | MCSS–Discrete |
| VH861697 | FT | Both | [no flag] | 4,386 | 0.52 | 0.57 | 2% | 1 | MCSS–Member |
| VH861815 | FT | 1 | [no flag] | 2,284 | 0.80 | 0.51 | 1% | 1 | MCSS–Member |
| VH861817 | FT | 2 | [no flag] | 2,120 | 0.75 | 0.55 | 1% | 1 | MCSS–Member |
| VH861822 | FT | Both | [no flag] | 4,403 | 0.84 | 0.59 | 1% | 1 | MCSS–Member |
| VH861827 | FT | 2 | [no flag] | 2,113 | 0.38 | 0.35 | 1% | 1 | MCSS–Member |
| VH861833 | FT | 1 | [no flag] | 2,283 | 0.56 | 0.48 | 1% | 1 | MCSS–Member |
| VH861848 | FT | Both | [no flag] | 4,393 | 0.42 | 0.39 | 2% | 1 | MCSS–Member |
| VH861861 | FT | 2 | [no flag] | 2,114 | 1.30 | 0.60 | 2% | 2 | Composite Objective–Member |
| VH862179 | FT | 1 | [no flag] | 2,285 | 0.78 | 0.49 | 1% | 1 | MCSS–Discrete |
| VH862191 | FT | 2 | [no flag] | 2,115 | 0.84 | 0.67 | 1% | 1 | MCSS–Discrete |
| VH862203 | FT | Both | [no flag] | 4,404 | 0.87 | 0.60 | 1% | 1 | MCSS–Discrete |
| VH862215 | FT | 2 | [no flag] | 2,116 | 0.64 | 0.60 | 2% | 1 | MCSS–Discrete |
| VH862224 | FT | 1 | [no flag] | 2,279 | 0.69 | 0.64 | 1% | 1 | MCSS–Discrete |
| VH862241 | FT | Both | [no flag] | 4,396 | 0.40 | 0.54 | 3% | 1 | MCSS–Discrete |
| VH862255 | FT | 2 | [no flag] | 2,116 | 1.35 | 0.69 | 2% | 2 | MCMS–Discrete |
| VH862268 | FT | Both | [no flag] | 4,383 | 0.65 | 0.43 | 1% | 1 | MCSS–Member |
| VH862275 | FT | 1 | [no flag] | 2,276 | 0.87 | 0.70 | 1% | 1 | MCSS–Member |
| VH862282 | FT | 2 | [no flag] | 2,108 | 0.78 | 0.65 | 1% | 1 | MCSS–Member |
| VH862379 | FT | Both | [no flag] | 4,388 | 0.72 | 0.54 | 2% | 1 | MCSS–Discrete |
| VH863175 | FT | 1 | L | 2,279 | 1.55 | 0.74 | 3% | 2 | GridMS–Discrete |
| VH863179 | FT | 2 | [no flag] | 2,110 | 1.54 | 0.70 | 3% | 2 | GridMS–Discrete |
| VH863922 | FT | Both | [no flag] | 4,382 | 0.66 | 0.71 | 5% | 1 | GridSS–Discrete |
| VH864008 | FT | 1 | [no flag] | 2,286 | 0.88 | 0.66 | 2% | 1 | MCSS–Discrete |
| VH864011 | FT | 2 | [no flag] | 2,119 | 0.81 | 0.58 | 1% | 1 | MCSS–Discrete |
| VH864019 | FT | 1 | [no flag] | 2,275 | 0.76 | 0.68 | 1% | 1 | MCSS–Member |
| VH864021 | FT | Both | [no flag] | 4,396 | 0.83 | 0.69 | 2% | 1 | MCSS–Discrete |
| VH864027 | FT | 2 | [no flag] | 2,115 | 0.74 | 0.66 | 2% | 1 | MCSS–Member |
| VH864031 | FT | 1 | [no flag] | 2,276 | 1.39 | 0.55 | 2% | 2 | MCMS–Discrete |
| VH864033 | FT | 2 | [no flag] | 2,117 | 1.31 | 0.60 | 2% | 2 | MCMS–Discrete |
| VH864048 | FT | Both | [no flag] | 4,395 | 0.86 | 0.72 | 2% | 1 | MCSS–Member |
| VH864068 | FT | Both | [no flag] | 4,384 | 0.80 | 0.60 | 2% | 1 | MCSS–Member |
| VH864074 | FT | Both | [no flag] | 4,384 | 0.60 | 0.60 | 2% | 1 | MCSS–Member |
| VH864080 | FT | 2 | [no flag] | 2,114 | 0.81 | 0.74 | 2% | 1 | MCSS–Member |
| VH864097 | FT | 1 | [no flag] | 2,278 | 0.52 | 0.35 | 3% | 1 | MatchSS–Discrete |
| VH864099 | FT | 1 | L | 2,280 | 0.79 | 0.42 | 3% | 2 | MatchMS–Discrete |
| VH877071 | FT | 1 | [no flag] | 2,280 | 1.46 | 0.66 | 3% | 2 | MatchMS–Discrete |
| VH882757 | FT | Both | P | 4,391 | 0.38 | 0.24 | 2% | 1 | MCSS–Discrete |
| VH891305 | FT | Both | [no flag] | 4,374 | 0.91 | 0.66 | 1% | 1 | MCSS–Member |
| VH891442 | FT | 2 | [no flag] | 2,117 | 0.97 | 0.42 | 2% | 2 | MCMS–Member |
| VH905647 | FT | Both | [no flag] | 4,391 | 0.98 | 0.49 | 4% | 2 | Composite Objective–Discrete |
| VH905759 | FT | 1 | [no flag] | 2,280 | 1.32 | 0.62 | 1% | 2 | Composite Objective–Member |

Table 7.A.3 Average Item Score and Polyserial Correlation for Each Item, Grade Eight

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Item ID | Item Use | Version | Flag | Sample Size | AIS | Polyserial | Omit Rate | Maximum Score Points | Item Type |
| VH871924 | FT | 2 | [no flag] | 2,161 | 0.92 | 0.65 | 1% | 1 | MCSS–Discrete |
| VH871926 | FT | 1 | [no flag] | 2,353 | 0.49 | 0.44 | 1% | 1 | MCSS–Discrete |
| VH871930 | FT | 2 | [no flag] | 2,156 | 0.37 | 0.40 | 1% | 1 | MCSS–Discrete |
| VH872139 | FT | Both | [no flag] | 4,503 | 0.56 | 0.39 | 2% | 1 | MCSS–Discrete |
| VH872143 | FT | 2 | [no flag] | 2,155 | 1.36 | 0.63 | 2% | 2 | MatchMS–Discrete |
| VH872146 | FT | 1 | [no flag] | 2,343 | 1.27 | 0.54 | 2% | 2 | MatchMS–Discrete |
| VH872197 | FT | 1 | [no flag] | 2,341 | 0.88 | 0.63 | 1% | 1 | MCSS–Member |
| VH872206 | FT | 2 | [no flag] | 2,154 | 0.83 | 0.62 | 1% | 1 | MCSS–Member |
| VH872213 | FT | Both | [no flag] | 4,496 | 0.93 | 0.64 | 1% | 1 | MCSS–Member |
| VH872216 | FT | Both | [no flag] | 4,502 | 0.70 | 0.64 | 2% | 1 | MCSS–Discrete |
| VH872218 | FT | Both | [no flag] | 4,503 | 0.45 | 0.57 | 2% | 1 | MCSS–Discrete |
| VH872222 | FT | 2 | [no flag] | 2,156 | 1.29 | 0.66 | 1% | 2 | MCMS–Discrete |
| VH872685 | FT | Both | [no flag] | 4,486 | 0.47 | 0.56 | 2% | 1 | MCSS–Discrete |
| VH872690 | FT | 2 | A P Rpoly | 2,154 | 0.29 | 0.15 | 3% | 1 | MCSS–Discrete |
| VH872899 | FT | Both | [no flag] | 4,485 | 0.67 | 0.43 | 2% | 1 | MCSS–Member |
| VH872909 | FT | 1 | [no flag] | 2,341 | 0.58 | 0.58 | 1% | 1 | MCSS–Member |
| VH872925 | FT | Both | [no flag] | 4,504 | 0.74 | 0.56 | 2% | 1 | MCSS–Discrete |
| VH872937 | FT | Both | [no flag] | 4,504 | 0.74 | 0.54 | 2% | 1 | MCSS–Member |
| VH872941 | FT | Both | [no flag] | 4,513 | 0.80 | 0.66 | 2% | 1 | MCSS–Member |
| VH873004 | FT | Both | [no flag] | 4,501 | 0.53 | 0.64 | 2% | 1 | MCSS–Member |
| VH873010 | FT | Both | [no flag] | 4,496 | 1.33 | 0.62 | 2% | 2 | MCMS–Discrete |
| VH873930 | FT | Both | O | 4,500 | 0.86 | 0.45 | 7% | 1 | ZoneSS–Discrete |
| VH873935 | FT | 1 | [no flag] | 2,352 | 0.80 | 0.65 | 1% | 1 | MCSS–Member |
| VH873990 | FT | Both | [no flag] | 4,507 | 0.47 | 0.49 | 1% | 1 | MCSS–Discrete |
| VH873991 | FT | 2 | [no flag] | 2,160 | 0.77 | 0.66 | 2% | 1 | MCSS–Member |
| VH874021 | FT | Both | [no flag] | 4,503 | 0.38 | 0.45 | 2% | 1 | MCSS–Member |
| VH874484 | FT | 1 | [no flag] | 2,346 | 0.74 | 0.53 | 2% | 1 | MCSS–Member |
| VH874548 | FT | Both | [no flag] | 4,502 | 0.64 | 0.73 | 1% | 1 | MCSS–Discrete |
| VH874586 | FT | Both | [no flag] | 4,501 | 0.58 | 0.50 | 2% | 1 | MCSS–Member |
| VH874606 | FT | Both | [no flag] | 4,502 | 0.65 | 0.70 | 2% | 1 | MCSS–Discrete |
| VH874610 | FT | 2 | [no flag] | 2,157 | 0.67 | 0.59 | 2% | 1 | MCSS–Discrete |
| VH874612 | FT | 2 | [no flag] | 2,155 | 1.25 | 0.52 | 3% | 2 | MCMS–Discrete |
| VH874648 | FT | 1 | [no flag] | 2,349 | 0.83 | 0.72 | 3% | 2 | Composite Objective–Discrete |
| VH877748 | FT | Both | [no flag] | 4,500 | 1.42 | 0.69 | 2% | 2 | MCMS–Discrete |
| VH887937 | FT | 1 | [no flag] | 2,340 | 0.92 | 0.60 | 2% | 1 | MCSS–Member |
| VH887965 | FT | Both | [no flag] | 4,502 | 0.83 | 0.63 | 1% | 1 | MCSS–Member |
| VH906457 | FT | 1 | [no flag] | 2,347 | 1.15 | 0.56 | 2% | 2 | Composite Objective–Member |
| VH906746 | FT | Both | [no flag] | 4,497 | 0.86 | 0.50 | 2% | 2 | Composite Objective–Member |
| VH914368 | FT | Both | [no flag] | 4,487 | 0.78 | 0.67 | 2% | 1 | MCSS–Discrete |

Table 7.A.4 Average Item Score and Polyserial Correlation for Each Item, High School

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Item ID | Item Use | Version | Flag | Sample Size | AIS | Polyserial | Omit Rate | Maximum Score Points | Item Type |
| VH861284 | FT | Both | [no flag] | 7,249 | 0.65 | 0.43 | 1% | 1 | MCSS–Member |
| VH861314 | FT | Both | [no flag] | 7,238 | 0.72 | 0.63 | 2% | 1 | MCSS–Member |
| VH861317 | FT | Both | [no flag] | 7,228 | 0.51 | 0.49 | 3% | 1 | MCSS–Member |
| VH861594 | FT | 2 | [no flag] | 3,413 | 1.15 | 0.57 | 2% | 2 | Composite Objective–Discrete |
| VH861657 | FT | 1 | [no flag] | 3,806 | 0.81 | 0.46 | 5% | 2 | MatchMS–Member |
| VH861665 | FT | Both | A | 7,224 | 0.33 | 0.56 | 2% | 1 | MCSS–Discrete |
| VH861706 | FT | Both | [no flag] | 7,211 | 0.82 | 0.64 | 2% | 1 | MCSS–Discrete |
| VH861715 | FT | 1 | [no flag] | 3,801 | 0.73 | 0.68 | 2% | 1 | MCSS–Member |
| VH861763 | FT | Both | [no flag] | 7,216 | 0.85 | 0.68 | 2% | 1 | MCSS–Member |
| VH861767 | FT | Both | [no flag] | 7,216 | 1.34 | 0.68 | 5% | 2 | MatchMS–Member |
| VH861770 | FT | 2 | [no flag] | 3,416 | 0.70 | 0.64 | 3% | 1 | MCSS–Discrete |
| VH861882 | FT | Both | [no flag] | 7,228 | 0.80 | 0.57 | 3% | 1 | MCSS–Discrete |
| VH861891 | FT | 1 | [no flag] | 3,805 | 0.52 | 0.29 | 2% | 1 | MCSS–Discrete |
| VH861895 | FT | Both | [no flag] | 7,244 | 0.63 | 0.44 | 2% | 1 | MCSS–Discrete |
| VH861904 | FT | 2 | [no flag] | 3,415 | 0.74 | 0.59 | 3% | 1 | MCSS–Discrete |
| VH861926 | FT | 1 | [no flag] | 3,797 | 0.80 | 0.64 | 2% | 1 | MCSS–Discrete |
| VH861941 | FT | 2 | [no flag] | 3,411 | 0.75 | 0.62 | 2% | 1 | MCSS–Discrete |
| VH861966 | FT | 2 | [no flag] | 3,403 | 0.70 | 0.52 | 3% | 1 | MCSS–Discrete |
| VH861971 | FT | 2 | [no flag] | 3,407 | 0.66 | 0.61 | 3% | 1 | MCSS–Discrete |
| VH861982 | FT | Both | [no flag] | 7,204 | 0.79 | 0.63 | 2% | 1 | MCSS–Discrete |
| VH862037 | FT | Both | [no flag] | 7,215 | 0.66 | 0.70 | 2% | 1 | MCSS–Member |
| VH862061 | FT | Both | [no flag] | 7,211 | 1.18 | 0.62 | 2% | 2 | MCMS–Discrete |
| VH863057 | FT | 1 | [no flag] | 3,820 | 0.79 | 0.54 | 1% | 1 | MCSS–Discrete |
| VH863073 | FT | 1 | [no flag] | 3,813 | 0.76 | 0.65 | 2% | 1 | MCSS–Member |
| VH863075 | FT | 2 | [no flag] | 3,418 | 0.36 | 0.42 | 3% | 1 | MCSS–Member |
| VH863079 | FT | Both | [no flag] | 7,230 | 0.45 | 0.46 | 3% | 1 | MCSS–Member |
| VH863082 | FT | 1 | [no flag] | 3,809 | 0.88 | 0.50 | 3% | 2 | MCMS–Member |
| VH863087 | FT | 2 | [no flag] | 3,421 | 0.68 | 0.48 | 2% | 1 | MCSS–Member |
| VH863112 | FT | 2 | [no flag] | 3,418 | 1.14 | 0.56 | 3% | 2 | MCMS–Member |
| VH863116 | FT | 2 | [no flag] | 3,415 | 0.71 | 0.59 | 3% | 1 | MCSS–Member |
| VH863131 | FT | 1 | [no flag] | 3,804 | 0.55 | 0.48 | 2% | 1 | MCSS–Member |
| VH863164 | FT | 2 | [no flag] | 3,413 | 0.59 | 0.47 | 3% | 1 | MCSS–Member |
| VH863168 | FT | Both | [no flag] | 7,218 | 0.44 | 0.38 | 3% | 1 | MCSS–Member |
| VH863171 | FT | Both | [no flag] | 7,215 | 0.52 | 0.32 | 3% | 1 | MCSS–Member |
| VH863173 | FT | 1 | [no flag] | 3,803 | 0.66 | 0.34 | 2% | 1 | MCSS–Member |
| VH877096 | FT | 1 | [no flag] | 3,796 | 0.80 | 0.68 | 2% | 1 | MCSS–Discrete |
| VH898133 | FT | Both | [no flag] | 7,212 | 1.24 | 0.66 | 2% | 2 | MCMS–Discrete |
| VH907952 | FT | Both | [no flag] | 7,215 | 1.32 | 0.63 | 2% | 2 | Composite Objective–Member |
| VH908011 | FT | 1 | [no flag] | 3,802 | 0.50 | 0.22 | 2% | 1 | MCSS–Member |
| VH917976 | FT | Both | [no flag] | 7,209 | 0.90 | 0.69 | 2% | 1 | MCSS–Discrete |
| VH918124 | FT | Both | [no flag] | 7,249 | 0.75 | 0.59 | 1% | 1 | MCSS–Member |

Table 7.A.5 Distribution of Item Scores for Each Item, Grade Five

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Item ID | Item Use | Versions | Maximum Points | Number of Students | Score 0 | Score 1 | Score 2 | Blank |
| VH861679 | FT | 1 | 1 | 2,279 | 59% | 40% | N/A | 1% |
| VH861685 | FT | 2 | 1 | 2,108 | 20% | 78% | N/A | 2% |
| VH861690 | FT | Both | 1 | 4,388 | 31% | 67% | N/A | 2% |
| VH861697 | FT | Both | 1 | 4,386 | 47% | 52% | N/A | 2% |
| VH861815 | FT | 1 | 1 | 2,284 | 19% | 80% | N/A | 1% |
| VH861817 | FT | 2 | 1 | 2,120 | 24% | 75% | N/A | 1% |
| VH861822 | FT | Both | 1 | 4,403 | 15% | 84% | N/A | 1% |
| VH861827 | FT | 2 | 1 | 2,113 | 60% | 38% | N/A | 1% |
| VH861833 | FT | 1 | 1 | 2,283 | 43% | 56% | N/A | 1% |
| VH861848 | FT | Both | 1 | 4,393 | 57% | 42% | N/A | 2% |
| VH861861 | FT | 2 | 2 | 2,114 | 17% | 32% | 49% | 2% |
| VH862179 | FT | 1 | 1 | 2,285 | 21% | 78% | N/A | 1% |
| VH862191 | FT | 2 | 1 | 2,115 | 15% | 84% | N/A | 1% |
| VH862203 | FT | Both | 1 | 4,404 | 12% | 87% | N/A | 1% |
| VH862215 | FT | 2 | 1 | 2,116 | 34% | 64% | N/A | 2% |
| VH862224 | FT | 1 | 1 | 2,279 | 29% | 69% | N/A | 1% |
| VH862241 | FT | Both | 1 | 4,396 | 58% | 40% | N/A | 3% |
| VH862255 | FT | 2 | 2 | 2,116 | 8% | 45% | 45% | 2% |
| VH862268 | FT | Both | 1 | 4,383 | 34% | 65% | N/A | 1% |
| VH862275 | FT | 1 | 1 | 2,276 | 12% | 87% | N/A | 1% |
| VH862282 | FT | 2 | 1 | 2,108 | 21% | 78% | N/A | 1% |
| VH862379 | FT | Both | 1 | 4,388 | 26% | 72% | N/A | 2% |
| VH863175 | FT | 1 | 2 | 2,279 | 2% | 35% | 60% | 3% |
| VH863179 | FT | 2 | 2 | 2,110 | 4% | 33% | 60% | 3% |
| VH863922 | FT | Both | 1 | 4,382 | 30% | 66% | N/A | 5% |
| VH864008 | FT | 1 | 1 | 2,286 | 10% | 88% | N/A | 2% |
| VH864011 | FT | 2 | 1 | 2,119 | 18% | 81% | N/A | 1% |
| VH864019 | FT | 1 | 1 | 2,275 | 23% | 76% | N/A | 1% |
| VH864021 | FT | Both | 1 | 4,396 | 15% | 83% | N/A | 2% |
| VH864027 | FT | 2 | 1 | 2,115 | 25% | 74% | N/A | 2% |
| VH864031 | FT | 1 | 2 | 2,276 | 6% | 46% | 47% | 2% |
| VH864033 | FT | 2 | 2 | 2,117 | 10% | 45% | 43% | 2% |
| VH864048 | FT | Both | 1 | 4,395 | 13% | 86% | N/A | 2% |
| VH864068 | FT | Both | 1 | 4,384 | 19% | 80% | N/A | 2% |
| VH864074 | FT | Both | 1 | 4,384 | 39% | 60% | N/A | 2% |
| VH864080 | FT | 2 | 1 | 2,114 | 17% | 81% | N/A | 2% |
| VH864097 | FT | 1 | 1 | 2,278 | 45% | 52% | N/A | 3% |
| VH864099 | FT | 1 | 2 | 2,280 | 58% | 0% | 39% | 3% |
| VH877071 | FT | 1 | 2 | 2,280 | 6% | 36% | 55% | 3% |
| VH882757 | FT | Both | 1 | 4,391 | 60% | 38% | N/A | 2% |
| VH891305 | FT | Both | 1 | 4,374 | 7% | 91% | N/A | 1% |
| VH891442 | FT | 2 | 2 | 2,117 | 22% | 55% | 21% | 2% |
| VH905647 | FT | Both | 2 | 4,391 | 24% | 47% | 25% | 4% |
| VH905759 | FT | 1 | 2 | 2,280 | 13% | 41% | 45% | 1% |

Table 7.A.6 Distribution of Item Scores for Each Item, Grade Eight

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Item ID | Item Use | Versions | Maximum Points | Number of Students | Score 0 | Score 1 | Score 2 | Blank |
| VH871924 | FT | 2 | 1 | 2,161 | 7% | 92% | N/A | 1% |
| VH871926 | FT | 1 | 1 | 2,353 | 50% | 49% | N/A | 1% |
| VH871930 | FT | 2 | 1 | 2,156 | 61% | 37% | N/A | 1% |
| VH872139 | FT | Both | 1 | 4,503 | 43% | 56% | N/A | 2% |
| VH872143 | FT | 2 | 2 | 2,155 | 10% | 39% | 48% | 2% |
| VH872146 | FT | 1 | 2 | 2,343 | 11% | 48% | 40% | 2% |
| VH872197 | FT | 1 | 1 | 2,341 | 11% | 88% | N/A | 1% |
| VH872206 | FT | 2 | 1 | 2,154 | 15% | 83% | N/A | 1% |
| VH872213 | FT | Both | 1 | 4,496 | 6% | 93% | N/A | 1% |
| VH872216 | FT | Both | 1 | 4,502 | 28% | 70% | N/A | 2% |
| VH872218 | FT | Both | 1 | 4,503 | 53% | 45% | N/A | 2% |
| VH872222 | FT | 2 | 2 | 2,156 | 16% | 37% | 46% | 1% |
| VH872685 | FT | Both | 1 | 4,486 | 51% | 47% | N/A | 2% |
| VH872690 | FT | 2 | 1 | 2,154 | 68% | 29% | N/A | 3% |
| VH872899 | FT | Both | 1 | 4,485 | 31% | 67% | N/A | 2% |
| VH872909 | FT | 1 | 1 | 2,341 | 41% | 58% | N/A | 1% |
| VH872925 | FT | Both | 1 | 4,504 | 24% | 74% | N/A | 2% |
| VH872937 | FT | Both | 1 | 4,504 | 24% | 74% | N/A | 2% |
| VH872941 | FT | Both | 1 | 4,513 | 19% | 80% | N/A | 2% |
| VH873004 | FT | Both | 1 | 4,501 | 45% | 53% | N/A | 2% |
| VH873010 | FT | Both | 2 | 4,496 | 9% | 45% | 44% | 2% |
| VH873930 | FT | Both | 1 | 4,500 | 7% | 86% | N/A | 7% |
| VH873935 | FT | 1 | 1 | 2,352 | 18% | 80% | N/A | 1% |
| VH873990 | FT | Both | 1 | 4,507 | 52% | 47% | N/A | 1% |
| VH873991 | FT | 2 | 1 | 2,160 | 22% | 77% | N/A | 2% |
| VH874021 | FT | Both | 1 | 4,503 | 60% | 38% | N/A | 2% |
| VH874484 | FT | 1 | 1 | 2,346 | 24% | 74% | N/A | 2% |
| VH874548 | FT | Both | 1 | 4,502 | 34% | 64% | N/A | 1% |
| VH874586 | FT | Both | 1 | 4,501 | 41% | 58% | N/A | 2% |
| VH874606 | FT | Both | 1 | 4,502 | 33% | 65% | N/A | 2% |
| VH874610 | FT | 2 | 1 | 2,157 | 31% | 67% | N/A | 2% |
| VH874612 | FT | 2 | 2 | 2,155 | 9% | 51% | 37% | 3% |
| VH874648 | FT | 1 | 2 | 2,349 | 47% | 17% | 33% | 3% |
| VH877748 | FT | Both | 2 | 4,500 | 6% | 43% | 49% | 2% |
| VH887937 | FT | 1 | 1 | 2,340 | 7% | 92% | N/A | 2% |
| VH887965 | FT | Both | 1 | 4,502 | 15% | 83% | N/A | 1% |
| VH906457 | FT | 1 | 2 | 2,347 | 17% | 49% | 33% | 2% |
| VH906746 | FT | Both | 2 | 4,497 | 32% | 47% | 19% | 2% |
| VH914368 | FT | Both | 1 | 4,487 | 20% | 78% | N/A | 2% |

Table 7.A.7 Distribution of Item Scores for Each Item, High School

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Item ID | Item Use | Versions | Maximum Points | Number of Students | Score 0 | Score 1 | Score 2 | Blank |
| VH861284 | FT | Both | 1 | 7,249 | 34% | 65% | N/A | 1% |
| VH861314 | FT | Both | 1 | 7,238 | 26% | 72% | N/A | 2% |
| VH861317 | FT | Both | 1 | 7,228 | 46% | 51% | N/A | 3% |
| VH861594 | FT | 2 | 2 | 3,413 | 20% | 41% | 37% | 2% |
| VH861657 | FT | 1 | 2 | 3,806 | 34% | 41% | 20% | 5% |
| VH861665 | FT | Both | 1 | 7,224 | 65% | 33% | N/A | 2% |
| VH861706 | FT | Both | 1 | 7,211 | 16% | 82% | N/A | 2% |
| VH861715 | FT | 1 | 1 | 3,801 | 25% | 73% | N/A | 2% |
| VH861763 | FT | Both | 1 | 7,216 | 14% | 85% | N/A | 2% |
| VH861767 | FT | Both | 2 | 7,216 | 12% | 33% | 51% | 5% |
| VH861770 | FT | 2 | 1 | 3,416 | 27% | 70% | N/A | 3% |
| VH861882 | FT | Both | 1 | 7,228 | 17% | 80% | N/A | 3% |
| VH861891 | FT | 1 | 1 | 3,805 | 46% | 52% | N/A | 2% |
| VH861895 | FT | Both | 1 | 7,244 | 35% | 63% | N/A | 2% |
| VH861904 | FT | 2 | 1 | 3,415 | 23% | 74% | N/A | 3% |
| VH861926 | FT | 1 | 1 | 3,797 | 18% | 80% | N/A | 2% |
| VH861941 | FT | 2 | 1 | 3,411 | 23% | 75% | N/A | 2% |
| VH861966 | FT | 2 | 1 | 3,403 | 27% | 70% | N/A | 3% |
| VH861971 | FT | 2 | 1 | 3,407 | 31% | 66% | N/A | 3% |
| VH861982 | FT | Both | 1 | 7,204 | 20% | 79% | N/A | 2% |
| VH862037 | FT | Both | 1 | 7,215 | 32% | 66% | N/A | 2% |
| VH862061 | FT | Both | 2 | 7,211 | 18% | 43% | 37% | 2% |
| VH863057 | FT | 1 | 1 | 3,820 | 20% | 79% | N/A | 1% |
| VH863073 | FT | 1 | 1 | 3,813 | 23% | 76% | N/A | 2% |
| VH863075 | FT | 2 | 1 | 3,418 | 61% | 36% | N/A | 3% |
| VH863079 | FT | Both | 1 | 7,230 | 51% | 45% | N/A | 3% |
| VH863082 | FT | 1 | 2 | 3,809 | 27% | 53% | 18% | 3% |
| VH863087 | FT | 2 | 1 | 3,421 | 30% | 68% | N/A | 2% |
| VH863112 | FT | 2 | 2 | 3,418 | 12% | 56% | 29% | 3% |
| VH863116 | FT | 2 | 1 | 3,415 | 26% | 71% | N/A | 3% |
| VH863131 | FT | 1 | 1 | 3,804 | 43% | 55% | N/A | 2% |
| VH863164 | FT | 2 | 1 | 3,413 | 38% | 59% | N/A | 3% |
| VH863168 | FT | Both | 1 | 7,218 | 53% | 44% | N/A | 3% |
| VH863171 | FT | Both | 1 | 7,215 | 45% | 52% | N/A | 3% |
| VH863173 | FT | 1 | 1 | 3,803 | 33% | 66% | N/A | 2% |
| VH877096 | FT | 1 | 1 | 3,796 | 19% | 80% | N/A | 2% |
| VH898133 | FT | Both | 2 | 7,212 | 8% | 55% | 34% | 2% |
| VH907952 | FT | Both | 2 | 7,215 | 14% | 36% | 48% | 2% |
| VH908011 | FT | 1 | 1 | 3,802 | 47% | 50% | N/A | 2% |
| VH917976 | FT | Both | 1 | 7,209 | 9% | 90% | N/A | 2% |
| VH918124 | FT | Both | 1 | 7,249 | 24% | 75% | N/A | 1% |

### Appendix 7.B: Completion Rates

**Note:** The full population was used.

Table 7.B.1 Percentage of Students in Each Grade Completing Embedded PTs

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Grade | No PTs Completed | 1 PT Completed | 2 PTs Completed | 3 PTs Completed | Number of Students Registered |
| Grade 5 | 10% | 2% | 0% | 88% | 5,847 |
| Grade 8 | 11% | 2% | 1% | 87% | 6,000 |
| High school | 25% | 1% | 1% | 73% | 12,133 |
| Grade 10 | 26% | 1% | 0% | 72% | 521 |
| Grade 11 | 21% | 1% | 1% | 77% | 3,471 |
| Grade 12 | 27% | 1% | 1% | 71% | 8,141 |

Table 7.B.2 Completion Rates by Grade Level for Each Embedded Performance Task (PT)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Grade Level | Embedded PT | Number of Students Registered | Number of Students Completing the PT | Percent Completing the PT |
| Grade 5 | Life Sciences | 5,847 | 5,259 | 90% |
| Grade 5 | Physical Sciences | 5,847 | 5,167 | 88% |
| Grade 5 | Earth and Space Sciences | 5,847 | 5,146 | 88% |
| Grade 8 | Life Sciences | 6,000 | 5,323 | 89% |
| Grade 8 | Physical Sciences | 6,000 | 5,254 | 88% |
| Grade 8 | Earth and Space Sciences | 6,000 | 5,239 | 87% |
| High school | Life Sciences | 12,133 | 8,891 | 73% |
| High school | Physical Sciences | 12,133 | 9,030 | 74% |
| High school | Earth and Space Sciences | 12,133 | 8,846 | 73% |
| Grade 10 | Life Sciences | 521 | 378 | 73% |
| Grade 10 | Physical Sciences | 521 | 384 | 74% |
| Grade 10 | Earth and Space Sciences | 521 | 378 | 73% |
| Grade 11 | Life Sciences | 3,471 | 2,679 | 77% |
| Grade 11 | Physical Sciences | 3,471 | 2,716 | 78% |
| Grade 11 | Earth and Space Sciences | 3,471 | 2,671 | 77% |
| Grade 12 | Life Sciences | 8,141 | 5,834 | 72% |
| Grade 12 | Physical Sciences | 8,141 | 5,930 | 73% |
| Grade 12 | Earth and Space Sciences | 8,141 | 5,797 | 71% |

### Appendix 7.C: Distribution of Raw Scores: Total Score Each Embedded PT

Table 7.C.1 Distribution of Total Score and PT Scores, Grade Five Version One

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 263 | 10% | 310 | 12% | 328 | 12% | 326 | 12% |
| 1 | 15 | 1% | 22 | 1% | 24 | 1% | 31 | 1% |
| 2 | 9 | 0% | 28 | 1% | 27 | 1% | 33 | 1% |
| 3 | 13 | 0% | 51 | 2% | 55 | 2% | 104 | 4% |
| 4 | 13 | 0% | 113 | 4% | 86 | 3% | 222 | 8% |
| 5 | 18 | 1% | 130 | 5% | 161 | 6% | 309 | 12% |
| 6 | 13 | 0% | 190 | 7% | 207 | 8% | 356 | 13% |
| 7 | 13 | 0% | 264 | 10% | 226 | 9% | 348 | 13% |
| 8 | 15 | 1% | 350 | 13% | 244 | 9% | 320 | 12% |
| 9 | 18 | 1% | 377 | 14% | 247 | 9% | 268 | 10% |
| 10 | 14 | 1% | 394 | 15% | 343 | 13% | 202 | 8% |
| 11 | 24 | 1% | 222 | 8% | 377 | 14% | 81 | 3% |
| 12 | 30 | 1% | 203 | 8% | 329 | 12% | 54 | 2% |
| 13 | 49 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 52 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 58 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 71 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 81 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 96 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 85 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 100 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 106 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 97 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 130 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 128 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 127 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 137 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 129 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 126 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 139 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 102 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 122 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 115 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 67 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 42 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 21 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 16 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.2 Distribution of Total Score and PT Scores, Grade Five Version Two

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 225 | 9% | 265 | 11% | 292 | 12% | 311 | 13% |
| 1 | 35 | 1% | 40 | 2% | 28 | 1% | 36 | 1% |
| 2 | 10 | 0% | 27 | 1% | 24 | 1% | 52 | 2% |
| 3 | 16 | 1% | 53 | 2% | 54 | 2% | 105 | 4% |
| 4 | 7 | 0% | 79 | 3% | 110 | 4% | 168 | 7% |
| 5 | 12 | 0% | 151 | 6% | 178 | 7% | 246 | 10% |
| 6 | 9 | 0% | 166 | 7% | 219 | 9% | 342 | 14% |
| 7 | 16 | 1% | 216 | 9% | 233 | 9% | 308 | 12% |
| 8 | 16 | 1% | 244 | 10% | 212 | 9% | 322 | 13% |
| 9 | 17 | 1% | 324 | 13% | 254 | 10% | 234 | 9% |
| 10 | 18 | 1% | 343 | 14% | 284 | 11% | 178 | 7% |
| 11 | 18 | 1% | 366 | 15% | 307 | 12% | 117 | 5% |
| 12 | 40 | 2% | 202 | 8% | 281 | 11% | 57 | 2% |
| 13 | 31 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 56 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 50 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 73 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 80 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 79 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 93 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 96 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 107 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 115 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 106 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 97 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 113 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 109 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 107 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 106 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 117 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 100 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 117 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 94 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 77 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 48 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 44 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 22 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.3 Distribution of Total Score and PT Scores, Grade Eight Version One

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 214 | 8% | 242 | 9% | 290 | 11% | 270 | 10% |
| 1 | 14 | 1% | 18 | 1% | 21 | 1% | 30 | 1% |
| 2 | 10 | 0% | 26 | 1% | 31 | 1% | 59 | 2% |
| 3 | 18 | 1% | 49 | 2% | 102 | 4% | 143 | 5% |
| 4 | 10 | 0% | 72 | 3% | 181 | 7% | 244 | 9% |
| 5 | 5 | 0% | 164 | 6% | 255 | 10% | 261 | 10% |
| 6 | 9 | 0% | 280 | 10% | 311 | 12% | 250 | 9% |
| 7 | 10 | 0% | 371 | 14% | 285 | 11% | 242 | 9% |
| 8 | 18 | 1% | 366 | 14% | 273 | 10% | 217 | 8% |
| 9 | 11 | 0% | 362 | 13% | 271 | 10% | 247 | 9% |
| 10 | 29 | 1% | 306 | 11% | 261 | 10% | 285 | 11% |
| 11 | 26 | 1% | 229 | 9% | 263 | 10% | 267 | 10% |
| 12 | 47 | 2% | 197 | 7% | 138 | 5% | 167 | 6% |
| 13 | 59 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 94 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 117 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 98 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 106 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 122 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 116 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 105 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 101 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 92 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 95 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 86 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 101 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 102 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 99 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 110 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 104 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 119 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 110 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 85 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 95 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 77 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 45 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 23 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.4 Distribution of Total Score and PT Scores, Grade Eight Version Two

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 250 | 10% | 294 | 12% | 340 | 13% | 309 | 12% |
| 1 | 31 | 1% | 24 | 1% | 31 | 1% | 47 | 2% |
| 2 | 13 | 1% | 33 | 1% | 61 | 2% | 31 | 1% |
| 3 | 15 | 1% | 53 | 2% | 99 | 4% | 75 | 3% |
| 4 | 8 | 0% | 94 | 4% | 175 | 7% | 179 | 7% |
| 5 | 16 | 1% | 181 | 7% | 264 | 10% | 255 | 10% |
| 6 | 16 | 1% | 229 | 9% | 278 | 11% | 248 | 10% |
| 7 | 13 | 1% | 288 | 11% | 302 | 12% | 249 | 10% |
| 8 | 15 | 1% | 305 | 12% | 285 | 11% | 241 | 10% |
| 9 | 15 | 1% | 281 | 11% | 279 | 11% | 291 | 11% |
| 10 | 19 | 1% | 311 | 12% | 247 | 10% | 266 | 10% |
| 11 | 25 | 1% | 257 | 10% | 135 | 5% | 227 | 9% |
| 12 | 47 | 2% | 184 | 7% | 38 | 1% | 116 | 5% |
| 13 | 49 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 70 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 75 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 93 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 111 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 117 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 106 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 118 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 101 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 94 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 87 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 88 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 97 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 95 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 99 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 121 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 97 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 97 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 110 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 83 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 75 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 48 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 16 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 4 | 0% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.5 Distribution of Total Score and PT Scores, High School Version One

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 671 | 14% | 793 | 17% | 811 | 17% | 783 | 17% |
| 1 | 45 | 1% | 49 | 1% | 51 | 1% | 70 | 1% |
| 2 | 21 | 0% | 58 | 1% | 49 | 1% | 82 | 2% |
| 3 | 25 | 1% | 116 | 2% | 115 | 2% | 158 | 3% |
| 4 | 33 | 1% | 223 | 5% | 183 | 4% | 279 | 6% |
| 5 | 27 | 1% | 417 | 9% | 320 | 7% | 477 | 10% |
| 6 | 33 | 1% | 460 | 10% | 437 | 9% | 578 | 12% |
| 7 | 22 | 0% | 519 | 11% | 546 | 12% | 603 | 13% |
| 8 | 23 | 0% | 521 | 11% | 562 | 12% | 544 | 12% |
| 9 | 26 | 1% | 498 | 11% | 517 | 11% | 428 | 9% |
| 10 | 33 | 1% | 500 | 11% | 458 | 10% | 314 | 7% |
| 11 | 43 | 1% | 342 | 7% | 401 | 9% | 266 | 6% |
| 12 | 69 | 1% | 210 | 4% | 256 | 5% | 124 | 3% |
| 13 | 65 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 112 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 138 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 149 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 176 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 203 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 231 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 206 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 245 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 207 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 174 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 199 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 192 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 176 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 172 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 171 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 165 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 126 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 126 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 146 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 107 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 86 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 50 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 13 | 0% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.6 Distribution of Total Score and PT Scores, High School Version Two

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 513 | 12% | 601 | 15% | 618 | 15% | 626 | 15% |
| 1 | 54 | 1% | 64 | 2% | 67 | 2% | 57 | 1% |
| 2 | 20 | 0% | 64 | 2% | 49 | 1% | 67 | 2% |
| 3 | 23 | 1% | 103 | 3% | 100 | 2% | 136 | 3% |
| 4 | 21 | 1% | 205 | 5% | 179 | 4% | 313 | 8% |
| 5 | 15 | 0% | 291 | 7% | 289 | 7% | 470 | 11% |
| 6 | 28 | 1% | 382 | 9% | 369 | 9% | 555 | 14% |
| 7 | 28 | 1% | 439 | 11% | 389 | 9% | 507 | 12% |
| 8 | 21 | 1% | 406 | 10% | 396 | 10% | 436 | 11% |
| 9 | 19 | 0% | 455 | 11% | 407 | 10% | 379 | 9% |
| 10 | 22 | 1% | 447 | 11% | 439 | 11% | 282 | 7% |
| 11 | 37 | 1% | 379 | 9% | 449 | 11% | 190 | 5% |
| 12 | 45 | 1% | 272 | 7% | 357 | 9% | 90 | 2% |
| 13 | 79 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 103 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 128 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 141 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 180 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 186 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 184 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 187 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 163 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 149 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 134 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 166 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 147 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 168 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 161 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 145 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 139 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 138 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 144 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 123 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 118 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 89 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 66 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 24 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.7 Distribution of Total Score and PT Scores, Grade Ten Version One

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 14 | 10% | 20 | 14% | 20 | 14% | 18 | 13% |
| 1 | 3 | 2% | 0 | 0% | 3 | 2% | 4 | 3% |
| 2 | 0 | 0% | 4 | 3% | 1 | 1% | 2 | 1% |
| 3 | 3 | 2% | 4 | 3% | 3 | 2% | 8 | 6% |
| 4 | 1 | 1% | 3 | 2% | 4 | 3% | 8 | 6% |
| 5 | 2 | 1% | 14 | 10% | 11 | 8% | 11 | 8% |
| 6 | 2 | 1% | 8 | 6% | 9 | 6% | 20 | 14% |
| 7 | 0 | 0% | 13 | 9% | 16 | 11% | 18 | 13% |
| 8 | 0 | 0% | 15 | 10% | 12 | 8% | 15 | 10% |
| 9 | 0 | 0% | 22 | 15% | 19 | 13% | 11 | 8% |
| 10 | 0 | 0% | 21 | 15% | 24 | 17% | 12 | 8% |
| 11 | 2 | 1% | 13 | 9% | 14 | 10% | 11 | 8% |
| 12 | 4 | 3% | 6 | 4% | 7 | 5% | 5 | 3% |
| 13 | 0 | 0% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 0 | 0% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 3 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 4 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 5 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 5 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 9 | 6% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 5 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 5 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 4 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 9 | 6% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 10 | 7% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 6 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 8 | 6% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 2 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 5 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 7 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 4 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 7 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 2 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 6 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 2 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 4 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 0 | 0% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.8 Distribution of Total Score and PT Scores, Grade Ten Version Two

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 27 | 12% | 36 | 15% | 32 | 14% | 36 | 15% |
| 1 | 6 | 3% | 6 | 3% | 8 | 3% | 6 | 3% |
| 2 | 1 | 0% | 0 | 0% | 7 | 3% | 5 | 2% |
| 3 | 1 | 0% | 5 | 2% | 7 | 3% | 7 | 3% |
| 4 | 0 | 0% | 6 | 3% | 9 | 4% | 12 | 5% |
| 5 | 2 | 1% | 20 | 9% | 13 | 6% | 27 | 12% |
| 6 | 6 | 3% | 25 | 11% | 18 | 8% | 36 | 15% |
| 7 | 2 | 1% | 26 | 11% | 21 | 9% | 29 | 12% |
| 8 | 0 | 0% | 27 | 12% | 23 | 10% | 22 | 9% |
| 9 | 2 | 1% | 22 | 9% | 22 | 9% | 22 | 9% |
| 10 | 1 | 0% | 17 | 7% | 25 | 11% | 13 | 6% |
| 11 | 2 | 1% | 26 | 11% | 25 | 11% | 14 | 6% |
| 12 | 2 | 1% | 18 | 8% | 24 | 10% | 5 | 2% |
| 13 | 5 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 3 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 5 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 7 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 12 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 8 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 17 | 7% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 10 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 5 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 6 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 12 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 6 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 10 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 6 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 14 | 6% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 9 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 6 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 7 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 6 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 4 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 10 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 8 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 5 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 1 | 0% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.9 Distribution of Total Score and PT Scores, Grade Eleven Version One

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 149 | 11% | 174 | 12% | 181 | 13% | 179 | 13% |
| 1 | 15 | 1% | 17 | 1% | 18 | 1% | 24 | 2% |
| 2 | 8 | 1% | 15 | 1% | 16 | 1% | 27 | 2% |
| 3 | 4 | 0% | 40 | 3% | 36 | 3% | 42 | 3% |
| 4 | 3 | 0% | 65 | 5% | 51 | 4% | 67 | 5% |
| 5 | 7 | 1% | 124 | 9% | 96 | 7% | 128 | 9% |
| 6 | 8 | 1% | 132 | 9% | 125 | 9% | 171 | 12% |
| 7 | 6 | 0% | 164 | 12% | 171 | 12% | 197 | 14% |
| 8 | 6 | 0% | 135 | 10% | 166 | 12% | 175 | 13% |
| 9 | 7 | 1% | 175 | 13% | 156 | 11% | 144 | 10% |
| 10 | 9 | 1% | 166 | 12% | 152 | 11% | 109 | 8% |
| 11 | 11 | 1% | 117 | 8% | 137 | 10% | 93 | 7% |
| 12 | 22 | 2% | 72 | 5% | 91 | 7% | 40 | 3% |
| 13 | 19 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 39 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 48 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 45 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 43 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 58 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 69 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 68 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 68 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 57 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 42 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 51 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 58 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 66 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 66 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 63 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 57 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 44 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 42 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 44 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 40 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 29 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 19 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 6 | 0% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.10 Distribution of Total Score and PT Scores, Grade Eleven Version Two

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 138 | 11% | 158 | 13% | 168 | 13% | 161 | 13% |
| 1 | 16 | 1% | 20 | 2% | 16 | 1% | 12 | 1% |
| 2 | 1 | 0% | 16 | 1% | 13 | 1% | 24 | 2% |
| 3 | 6 | 0% | 38 | 3% | 35 | 3% | 49 | 4% |
| 4 | 4 | 0% | 70 | 6% | 49 | 4% | 110 | 9% |
| 5 | 3 | 0% | 90 | 7% | 99 | 8% | 147 | 12% |
| 6 | 7 | 1% | 128 | 10% | 117 | 9% | 159 | 13% |
| 7 | 7 | 1% | 140 | 11% | 113 | 9% | 162 | 13% |
| 8 | 10 | 1% | 123 | 10% | 139 | 11% | 145 | 11% |
| 9 | 4 | 0% | 135 | 11% | 142 | 11% | 116 | 9% |
| 10 | 6 | 0% | 143 | 11% | 118 | 9% | 93 | 7% |
| 11 | 13 | 1% | 125 | 10% | 136 | 11% | 55 | 4% |
| 12 | 15 | 1% | 78 | 6% | 119 | 9% | 31 | 2% |
| 13 | 29 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 41 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 41 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 43 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 56 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 55 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 60 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 58 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 46 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 53 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 36 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 63 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 42 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 49 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 56 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 45 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 49 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 37 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 48 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 37 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 35 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 23 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 23 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 9 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.11 Distribution of Total Score and PT Scores, Grade Twelve Version One

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 508 | 16% | 599 | 19% | 610 | 19% | 586 | 19% |
| 1 | 27 | 1% | 32 | 1% | 30 | 1% | 42 | 1% |
| 2 | 13 | 0% | 39 | 1% | 32 | 1% | 53 | 2% |
| 3 | 18 | 1% | 72 | 2% | 76 | 2% | 108 | 3% |
| 4 | 29 | 1% | 155 | 5% | 128 | 4% | 204 | 6% |
| 5 | 18 | 1% | 279 | 9% | 213 | 7% | 338 | 11% |
| 6 | 23 | 1% | 320 | 10% | 303 | 10% | 387 | 12% |
| 7 | 16 | 1% | 342 | 11% | 359 | 11% | 388 | 12% |
| 8 | 17 | 1% | 371 | 12% | 384 | 12% | 354 | 11% |
| 9 | 19 | 1% | 301 | 10% | 342 | 11% | 273 | 9% |
| 10 | 24 | 1% | 313 | 10% | 282 | 9% | 193 | 6% |
| 11 | 30 | 1% | 212 | 7% | 250 | 8% | 162 | 5% |
| 12 | 43 | 1% | 132 | 4% | 158 | 5% | 79 | 2% |
| 13 | 46 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 73 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 87 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 100 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 128 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 140 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 153 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 133 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 172 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 146 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 123 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 138 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 128 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 102 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 104 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 103 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 101 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 78 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 77 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 100 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 61 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 55 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 27 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 7 | 0% | N/A | N/A | N/A | N/A | N/A | N/A |

Table 7.C.12 Distribution of Total Score and PT Scores, Grade Twelve Version Two

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Raw Score | Number of Students (Total) | Percentage of Students (Total) | Number of Students (PT 1, Life Sciences) | Percentage of Students (PT 1, Life Sciences) | Number of Students (PT 2, Physical Sciences) | Percentage of Students (PT 2, Physical Sciences) | Number of Students (PT 3, Earth and Space Sciences) | Percentage of Students (PT 3, Earth and Space Sciences) |
| 0 | 348 | 13% | 407 | 16% | 418 | 16% | 429 | 16% |
| 1 | 32 | 1% | 38 | 1% | 43 | 2% | 39 | 1% |
| 2 | 18 | 1% | 48 | 2% | 29 | 1% | 38 | 1% |
| 3 | 16 | 1% | 60 | 2% | 58 | 2% | 80 | 3% |
| 4 | 17 | 1% | 129 | 5% | 121 | 5% | 191 | 7% |
| 5 | 10 | 0% | 181 | 7% | 177 | 7% | 296 | 11% |
| 6 | 15 | 1% | 229 | 9% | 234 | 9% | 360 | 14% |
| 7 | 19 | 1% | 273 | 10% | 255 | 10% | 316 | 12% |
| 8 | 11 | 0% | 256 | 10% | 234 | 9% | 269 | 10% |
| 9 | 13 | 0% | 298 | 11% | 243 | 9% | 241 | 9% |
| 10 | 15 | 1% | 287 | 11% | 296 | 11% | 176 | 7% |
| 11 | 22 | 1% | 228 | 9% | 288 | 11% | 121 | 5% |
| 12 | 28 | 1% | 176 | 7% | 214 | 8% | 54 | 2% |
| 13 | 45 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 14 | 59 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 15 | 82 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 16 | 91 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 17 | 112 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 18 | 123 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 19 | 107 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 20 | 119 | 5% | N/A | N/A | N/A | N/A | N/A | N/A |
| 21 | 112 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 22 | 90 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 23 | 86 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 24 | 97 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 25 | 95 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 26 | 113 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 27 | 91 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 28 | 91 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 29 | 84 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 30 | 94 | 4% | N/A | N/A | N/A | N/A | N/A | N/A |
| 31 | 90 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 32 | 82 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 33 | 73 | 3% | N/A | N/A | N/A | N/A | N/A | N/A |
| 34 | 58 | 2% | N/A | N/A | N/A | N/A | N/A | N/A |
| 35 | 38 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |
| 36 | 14 | 1% | N/A | N/A | N/A | N/A | N/A | N/A |

### Appendix 7.D: DIF Analysis: Number and Percentage of Items in Each DIF Category

Table 7.D.1 Categorization of DIF, Grade Five

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DIF Category | Male–Female N | Male–Female Pct | White–African Amer N | White–African Amer Pct | White–Asian N | White–Asian Pct | White–Filipino N | White–Filipino Pct | White–Hispanic N | White–Hispanic Pct |
| C+ | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| B+ | 0 | 0% | 0 | 0% | 0 | 0% | 1 | 2% | 0 | 0% |
| A+ | 22 | 50% | 21 | 48% | 25 | 57% | 6 | 14% | 23 | 52% |
| A- | 22 | 50% | 21 | 48% | 18 | 41% | 4 | 9% | 20 | 45% |
| B- | 0 | 0% | 2 | 5% | 1 | 2% | 1 | 2% | 1 | 2% |
| C- | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| N/A | 0 | 0% | 0 | 0% | 0 | 0% | 32 | 73% | 0 | 0% |
| **Items Total** | **44** | **100%** | **44** | **100%** | **44** | **100%** | **44** | **100%** | **44** | **100%** |

Table 7.D.1 Categorization of DIF, Grade Five (Continued)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DIF Category | Intellectual Disability–Autism N | Intellectual Disability–Autism Pct | Intellectual Disability–Multiple Disabilities N | Intellectual Disability–Multiple Disabilities Pct | Intellectual Disability–Orthopedic Impairment N | Intellectual Disability–Orthopedic Impairment Pct | Intellectual Disability–Other N | Intellectual Disability–Other Pct | Intellectual Disability–Specific Learning N | Intellectual Disability–Specific Learning Pct | Intellectual Disability–Speech or Language N | Intellectual Disability–Speech or Language Pct |
| C+ | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 1 | 2% | 0 | 0% |
| B+ | 1 | 2% | 0 | 0% | 1 | 2% | 1 | 2% | 2 | 5% | 1 | 2% |
| A+ | 19 | 43% | 4 | 9% | 6 | 14% | 20 | 45% | 24 | 55% | 10 | 23% |
| A- | 23 | 52% | 10 | 23% | 7 | 16% | 23 | 52% | 15 | 34% | 5 | 11% |
| B- | 1 | 2% | 2 | 5% | 2 | 5% | 0 | 0% | 2 | 5% | 0 | 0% |
| C- | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| N/A | 0 | 0% | 28 | 64% | 28 | 64% | 0 | 0% | 0 | 0% | 28 | 64% |
| **Items Total** | **44** | **100%** | **44** | **100%** | **44** | **100%** | **44** | **100%** | **44** | **100%** | **44** | **100%** |

Table 7.D.2 Categorization of DIF, Grade Eight

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DIF Category | Male–Female N | Male–Female Pct | White–African Amer N | White–African Amer Pct | White–Asian N | White–Asian Pct | White–Filipino N | White–Filipino Pct | White–Hispanic N | White–Hispanic Pct |
| C+ | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| B+ | 0 | 0% | 1 | 3% | 0 | 0% | 1 | 3% | 0 | 0% |
| A+ | 19 | 49% | 15 | 38% | 19 | 49% | 10 | 26% | 19 | 49% |
| A- | 20 | 51% | 23 | 59% | 19 | 49% | 8 | 21% | 20 | 51% |
| B- | 0 | 0% | 0 | 0% | 1 | 3% | 2 | 5% | 0 | 0% |
| C- | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| N/A | 0 | 0% | 0 | 0% | 0 | 0% | 18 | 46% | 0 | 0% |
| **Items Total** | **39** | **100%** | **39** | **100%** | **39** | **100%** | **39** | **100%** | **39** | **100%** |

Table 7.D.2 Categorization of DIF, Grade Eight (Continued)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DIF Category | Intellectual Disability–Autism N | Intellectual Disability–Autism Pct | Intellectual Disability–Multiple Disabilities N | Intellectual Disability–Multiple Disabilities Pct | Intellectual Disability–Orthopedic Impairment N | Intellectual Disability–Orthopedic Impairment Pct | Intellectual Disability–Other N | Intellectual Disability–Other Pct | Intellectual Disability–Specific Learning N | Intellectual Disability–Specific Learning Pct | Intellectual Disability–Speech or Language N | Intellectual Disability–Speech or Language Pct |
| C+ | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| B+ | 0 | 0% | 1 | 3% | 2 | 5% | 0 | 0% | 1 | 3% | 0 | 0% |
| A+ | 21 | 54% | 8 | 21% | 10 | 26% | 20 | 51% | 20 | 51% | 0 | 0% |
| A- | 18 | 46% | 11 | 28% | 9 | 23% | 17 | 44% | 18 | 46% | 0 | 0% |
| B- | 0 | 0% | 1 | 3% | 0 | 0% | 1 | 3% | 0 | 0% | 0 | 0% |
| C- | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| N/A | 0 | 0% | 18 | 46% | 18 | 46% | 1 | 3% | 0 | 0% | 39 | 100% |
| **Items Total** | **39** | **100%** | **39** | **100%** | **39** | **100%** | **39** | **100%** | **39** | **100%** | **39** | **100%** |

Table 7.D.3 Categorization of DIF, High School

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DIF Category | Male–Female N | Male–Female Pct | White–African Amer N | White–African Amer Pct | White–Asian N | White–Asian Pct | White–Filipino N | White–Filipino Pct | White–Hispanic N | White–Hispanic Pct |
| C+ | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| B+ | 0 | 0% | 4 | 10% | 0 | 0% | 0 | 0% | 0 | 0% |
| A+ | 22 | 54% | 16 | 39% | 17 | 41% | 23 | 56% | 20 | 49% |
| A- | 19 | 46% | 20 | 49% | 23 | 56% | 15 | 37% | 21 | 51% |
| B- | 0 | 0% | 1 | 2% | 1 | 2% | 3 | 7% | 0 | 0% |
| C- | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| N/A | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| **Items Total** | **41** | **100%** | **41** | **100%** | **41** | **100%** | **41** | **100%** | **41** | **100%** |

Table 7.D.3 Categorization of DIF, High School (Continued One)

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| DIF Category | Intellectual Disability–Autism N | Intellectual Disability–Autism Pct | Intellectual Disability–Multiple Disabilities N | Intellectual Disability–Multiple Disabilities Pct | Intellectual Disability–Orthopedic Impairment N | Intellectual Disability–Orthopedic Impairment Pct | Intellectual Disability–Other N | Intellectual Disability–Other Pct | Intellectual Disability–Specific Learning N | Intellectual Disability–Specific Learning Pct | Intellectual Disability–Speech or Language N | Intellectual Disability–Speech or Language Pct |
| C+ | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| B+ | 1 | 2% | 1 | 2% | 2 | 5% | 1 | 2% | 2 | 5% | 0 | 0% |
| A+ | 18 | 44% | 15 | 37% | 18 | 44% | 23 | 56% | 19 | 46% | 0 | 0% |
| A- | 21 | 51% | 11 | 27% | 20 | 49% | 16 | 39% | 19 | 46% | 0 | 0% |
| B- | 1 | 2% | 3 | 7% | 1 | 2% | 1 | 2% | 1 | 2% | 0 | 0% |
| C- | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% | 0 | 0% |
| N/A | 0 | 0% | 11 | 27% | 0 | 0% | 0 | 0% | 0 | 0% | 41 | 100% |
| **Items Total** | **41** | **100%** | **41** | **100%** | **41** | **100%** | **41** | **100%** | **41** | **100%** | **41** | **100%** |

Table 7.D.3 Categorization of DIF, High School (Continued Two)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| DIF Category | Grade 10 N | Grade 10 Pct | Grade 12 N | Grade 12 Pct |
| C+ | 0 | 0% | 0 | 0% |
| B+ | 0 | 0% | 0 | 0% |
| A+ | 22 | 54% | 19 | 46% |
| A- | 18 | 44% | 22 | 54% |
| B- | 1 | 2% | 0 | 0% |
| C- | 0 | 0% | 0 | 0% |
| N/A | 0 | 0% | 0 | 0% |
| **Items Total** | **41** | **100%** | **41** | **100%** |

Table 7.D.4 Items Exhibiting Significant DIF, Grade Five

|  |  |  |
| --- | --- | --- |
| Item ID | Item Sequence | Intellectual Disability–Specific Learning MH D-‍DIF Value |
| VH864074 | 9 | 2.47 |

### Appendix 7.E: Item Response Theory (IRT) Analyses Results

Table 7.E.1 IRT Parameter Estimates for All CAA for Science Items

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade | Number of Items | Average of *b*‑value | SD *b*‑value | Minimum *b*‑value | Maximum *b*‑value | RMSEA |
| Grade 5 | 44 | -1.0116 | 0.8830 | -2.7204 | 0.5856 | 0.03 |
| Grade 8 | 39 | -0.9036 | 1.0003 | -3.0102 | 1.0439 | 0.03 |
| High school | 41 | -0.7655 | 0.7581 | -2.4949 | 0.8572 | 0.04 |

Table 7.E.2 IRT Item Difficulty, Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item ID | *b-*value | *b-*value SE | *d-*values | *d-*values SE |
| VH861679 | 0.5279 | 0.0512 | N/A | N/A |
| VH861685 | -1.5516 | 0.0630 | N/A | N/A |
| VH861690 | -0.8724 | 0.0404 | N/A | N/A |
| VH861697 | -0.0756 | 0.0386 | N/A | N/A |
| VH861815 | -1.6116 | 0.0613 | N/A | N/A |
| VH861817 | -1.3695 | 0.0589 | N/A | N/A |
| VH861822 | -1.9147 | 0.0499 | N/A | N/A |
| VH861827 | 0.5509 | 0.0532 | N/A | N/A |
| VH861833 | -0.2672 | 0.0514 | N/A | N/A |
| VH861848 | 0.4138 | 0.0386 | N/A | N/A |
| VH861861 | -0.7361 | 0.0395 | 0.4307 : -0.4307 | 0.0502 : 0.0502 |
| VH862179 | -1.4551 | 0.0587 | N/A | N/A |
| VH862191 | -2.0097 | 0.0717 | N/A | N/A |
| VH862203 | -2.2246 | 0.0534 | N/A | N/A |
| VH862215 | -0.7384 | 0.0550 | N/A | N/A |
| VH862224 | -0.9420 | 0.0568 | N/A | N/A |
| VH862241 | 0.5035 | 0.0388 | N/A | N/A |
| VH862255 | -1.0524 | 0.0482 | 1.1511 : -1.1511 | 0.0525 : 0.0525 |
| VH862268 | -0.7462 | 0.0394 | N/A | N/A |
| VH862275 | -2.1516 | 0.0735 | N/A | N/A |
| VH862282 | -1.5242 | 0.0631 | N/A | N/A |
| VH862379 | -1.1132 | 0.0419 | N/A | N/A |
| VH863175 | -1.5829 | 0.0578 | 1.0152 : -1.0152 | 0.0639 : 0.0639 |
| VH863179 | -1.5438 | 0.0538 | 0.8693 : -0.8693 | 0.0622 : 0.0622 |
| VH863922 | -0.7691 | 0.0416 | N/A | N/A |
| VH864008 | -2.2766 | 0.0760 | N/A | N/A |
| VH864011 | -1.7614 | 0.0639 | N/A | N/A |
| VH864019 | -1.3176 | 0.0619 | N/A | N/A |
| VH864021 | -1.8652 | 0.0507 | N/A | N/A |
| VH864027 | -1.2585 | 0.0616 | N/A | N/A |
| VH864031 | -1.1649 | 0.0503 | 1.2527 : -1.2527 | 0.0535 : 0.0535 |
| VH864033 | -0.8879 | 0.0450 | 1.0472 : -1.0472 | 0.0503 : 0.0503 |
| VH864048 | -2.1112 | 0.0547 | N/A | N/A |
| VH864068 | -1.6351 | 0.0464 | N/A | N/A |
| VH864074 | -0.4647 | 0.0397 | N/A | N/A |
| VH864080 | -1.7786 | 0.0716 | N/A | N/A |
| VH864097 | -0.0854 | 0.0498 | N/A | N/A |
| VH864099 | 0.3701 | 0.0322 | -4.6500 : 4.6500 | 0.3592 : 0.3592 |
| VH877071 | -1.2150 | 0.0479 | 0.8406 : -0.8406 | 0.0544 : 0.0544 |
| VH882757 | 0.5856 | 0.0395 | N/A | N/A |
| VH891305 | -2.7204 | 0.0646 | N/A | N/A |
| VH891442 | 0.0604 | 0.0415 | 1.2897 : -1.2897 | 0.0456 : 0.0456 |
| VH905647 | 0.0690 | 0.0302 | 0.9491 : -0.9491 | 0.0312 : 0.0312 |
| VH905759 | -0.7968 | 0.0430 | 0.8413 : -0.8413 | 0.0479 : 0.0479 |

Table 7.E.3 IRT Item Difficulty, Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item ID | *b-*value | *b-*value SE | *d-*values | *d-*values SE |
| VH871924 | -2.9114 | 0.0894 | N/A | N/A |
| VH871926 | 0.0960 | 0.0505 | N/A | N/A |
| VH871930 | 0.5714 | 0.0524 | N/A | N/A |
| VH872139 | -0.2999 | 0.0389 | N/A | N/A |
| VH872143 | -1.0305 | 0.0439 | 0.8579 : -0.8579 | 0.0509 : 0.0509 |
| VH872146 | -0.7729 | 0.0421 | 1.1759 : -1.1759 | 0.0467 : 0.0467 |
| VH872197 | -2.3178 | 0.0722 | N/A | N/A |
| VH872206 | -1.9626 | 0.0677 | N/A | N/A |
| VH872213 | -3.0102 | 0.0664 | N/A | N/A |
| VH872216 | -1.0715 | 0.0413 | N/A | N/A |
| VH872218 | 0.2208 | 0.0390 | N/A | N/A |
| VH872222 | -0.7903 | 0.0412 | 0.6361 : -0.6361 | 0.0490 : 0.0490 |
| VH872685 | 0.1469 | 0.0389 | N/A | N/A |
| VH872690 | 1.0439 | 0.0542 | N/A | N/A |
| VH872899 | -0.8680 | 0.0401 | N/A | N/A |
| VH872909 | -0.3394 | 0.0517 | N/A | N/A |
| VH872925 | -1.2805 | 0.0425 | N/A | N/A |
| VH872937 | -1.2730 | 0.0422 | N/A | N/A |
| VH872941 | -1.6763 | 0.0457 | N/A | N/A |
| VH873004 | -0.1659 | 0.0387 | N/A | N/A |
| VH873010 | -0.9743 | 0.0343 | 1.1133 : -1.1133 | 0.0351 : 0.0351 |
| VH873930 | -2.1830 | 0.0500 | N/A | N/A |
| VH873935 | -1.6588 | 0.0609 | N/A | N/A |
| VH873990 | 0.1382 | 0.0389 | N/A | N/A |
| VH873991 | -1.5109 | 0.0612 | N/A | N/A |
| VH874021 | 0.5898 | 0.0393 | N/A | N/A |
| VH874484 | -1.2635 | 0.0563 | N/A | N/A |
| VH874548 | -0.7226 | 0.0417 | N/A | N/A |
| VH874586 | -0.3888 | 0.0392 | N/A | N/A |
| VH874606 | -0.7445 | 0.0415 | N/A | N/A |
| VH874610 | -0.9006 | 0.0556 | N/A | N/A |
| VH874612 | -0.8272 | 0.0451 | 1.3005 : -1.3005 | 0.0492 : 0.0492 |
| VH874648 | 0.3739 | 0.0362 | -0.4773 : 0.4773 | 0.0563 : 0.0563 |
| VH877748 | -1.2849 | 0.0377 | 1.1985 : -1.1985 | 0.0396 : 0.0396 |
| VH887937 | -2.7552 | 0.0836 | N/A | N/A |
| VH887965 | -1.9235 | 0.0490 | N/A | N/A |
| VH906457 | -0.3776 | 0.0398 | 1.1079 : -1.1079 | 0.0436 : 0.0436 |
| VH906746 | 0.3862 | 0.0315 | 1.0418 : -1.0418 | 0.0311 : 0.0311 |
| VH914368 | -1.5206 | 0.0455 | N/A | N/A |

Table 7.E.4 IRT Item Difficulty, High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item ID | *b-*value | *b-*value SE | *d-*values | *d-*values SE |
| VH861284 | -0.7513 | 0.0304 | N/A | N/A |
| VH861314 | -1.1408 | 0.0322 | N/A | N/A |
| VH861317 | -0.0790 | 0.0295 | N/A | N/A |
| VH861594 | -0.3922 | 0.0306 | 0.6877 : -0.6877 | 0.0363 : 0.0363 |
| VH861657 | 0.4547 | 0.0290 | 0.7020 : -0.7020 | 0.0344 : 0.0344 |
| VH861665 | 0.8572 | 0.0312 | N/A | N/A |
| VH861706 | -1.7689 | 0.0382 | N/A | N/A |
| VH861715 | -1.1723 | 0.0457 | N/A | N/A |
| VH861763 | -1.9907 | 0.0409 | N/A | N/A |
| VH861767 | -0.8037 | 0.0233 | 0.4580 : -0.4580 | 0.0279 : 0.0279 |
| VH861770 | -1.0214 | 0.0454 | N/A | N/A |
| VH861882 | -1.6723 | 0.0353 | N/A | N/A |
| VH861891 | -0.0762 | 0.0376 | N/A | N/A |
| VH861895 | -0.6328 | 0.0299 | N/A | N/A |
| VH861904 | -1.2685 | 0.0455 | N/A | N/A |
| VH861926 | -1.6095 | 0.0487 | N/A | N/A |
| VH861941 | -1.3574 | 0.0476 | N/A | N/A |
| VH861966 | -1.0351 | 0.0443 | N/A | N/A |
| VH861971 | -0.7954 | 0.0437 | N/A | N/A |
| VH861982 | -1.5351 | 0.0355 | N/A | N/A |
| VH862037 | -0.8004 | 0.0319 | N/A | N/A |
| VH862061 | -0.4433 | 0.0233 | 0.7881 : -0.7881 | 0.0262 : 0.0262 |
| VH863057 | -1.5740 | 0.0470 | N/A | N/A |
| VH863073 | -1.3209 | 0.0468 | N/A | N/A |
| VH863075 | 0.6538 | 0.0411 | N/A | N/A |
| VH863079 | 0.2175 | 0.0296 | N/A | N/A |
| VH863082 | 0.3506 | 0.0318 | 1.1979 : -1.1979 | 0.0341 : 0.0341 |
| VH863087 | -0.9232 | 0.0423 | N/A | N/A |
| VH863112 | -0.4481 | 0.0337 | 1.3421 : -1.3421 | 0.0373 : 0.0373 |
| VH863116 | -1.1122 | 0.0455 | N/A | N/A |
| VH863131 | -0.2171 | 0.0395 | N/A | N/A |
| VH863164 | -0.4430 | 0.0414 | N/A | N/A |
| VH863168 | 0.3041 | 0.0302 | N/A | N/A |
| VH863171 | -0.0956 | 0.0297 | N/A | N/A |
| VH863173 | -0.7605 | 0.0397 | N/A | N/A |
| VH877096 | -1.5907 | 0.0492 | N/A | N/A |
| VH898133 | -0.7879 | 0.0272 | 1.4449 : -1.4449 | 0.0282 : 0.0282 |
| VH907952 | -0.7860 | 0.0235 | 0.5768 : -0.5768 | 0.0271 : 0.0271 |
| VH908011 | -0.0123 | 0.0379 | N/A | N/A |
| VH917976 | -2.4949 | 0.0457 | N/A | N/A |
| VH918124 | -1.3123 | 0.0330 | N/A | N/A |

Table 7.E.5 Item Difficulties and Omit Rate, Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item ID | Item Type | *p*-value | IRT *b*-value | Omit Rate |
| VH861679 | FT | 0.40 | 0.53 | 1% |
| VH861685 | FT | 0.78 | -1.55 | 2% |
| VH861690 | FT | 0.67 | -0.87 | 2% |
| VH861697 | FT | 0.52 | -0.08 | 2% |
| VH861815 | FT | 0.80 | -1.61 | 1% |
| VH861817 | FT | 0.75 | -1.37 | 1% |
| VH861822 | FT | 0.84 | -1.91 | 1% |
| VH861827 | FT | 0.38 | 0.55 | 1% |
| VH861833 | FT | 0.56 | -0.27 | 1% |
| VH861848 | FT | 0.42 | 0.41 | 2% |
| VH861861 | FT | 0.65 | -0.74 | 2% |
| VH862179 | FT | 0.78 | -1.46 | 1% |
| VH862191 | FT | 0.84 | -2.01 | 1% |
| VH862203 | FT | 0.87 | -2.22 | 1% |
| VH862215 | FT | 0.64 | -0.74 | 2% |
| VH862224 | FT | 0.69 | -0.94 | 1% |
| VH862241 | FT | 0.40 | 0.50 | 3% |
| VH862255 | FT | 0.67 | -1.05 | 2% |
| VH862268 | FT | 0.65 | -0.75 | 1% |
| VH862275 | FT | 0.87 | -2.15 | 1% |
| VH862282 | FT | 0.78 | -1.52 | 1% |
| VH862379 | FT | 0.72 | -1.11 | 2% |
| VH863175 | FT | 0.77 | -1.58 | 3% |
| VH863179 | FT | 0.77 | -1.54 | 3% |
| VH863922 | FT | 0.66 | -0.77 | 5% |
| VH864008 | FT | 0.88 | -2.28 | 2% |
| VH864011 | FT | 0.81 | -1.76 | 1% |
| VH864019 | FT | 0.76 | -1.32 | 1% |
| VH864021 | FT | 0.83 | -1.87 | 2% |
| VH864027 | FT | 0.74 | -1.26 | 2% |
| VH864031 | FT | 0.70 | -1.16 | 2% |
| VH864033 | FT | 0.65 | -0.89 | 2% |
| VH864048 | FT | 0.86 | -2.11 | 2% |
| VH864068 | FT | 0.80 | -1.64 | 2% |
| VH864074 | FT | 0.60 | -0.46 | 2% |
| VH864080 | FT | 0.81 | -1.78 | 2% |
| VH864097 | FT | 0.52 | -0.09 | 3% |
| VH864099 | FT | 0.39 | 0.37 | 3% |
| VH877071 | FT | 0.73 | -1.22 | 3% |
| VH882757 | FT | 0.38 | 0.59 | 2% |
| VH891305 | FT | 0.91 | -2.72 | 1% |
| VH891442 | FT | 0.49 | 0.06 | 2% |
| VH905647 | FT | 0.49 | 0.07 | 4% |
| VH905759 | FT | 0.66 | -0.80 | 1% |

Table 7.E.6 Item Difficulties and Omit Rate, Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item ID | Item Type | *p*-value | IRT *b*-value | Omit Rate |
| VH871924 | FT | 0.92 | -2.91 | 1% |
| VH871926 | FT | 0.49 | 0.10 | 1% |
| VH871930 | FT | 0.37 | 0.57 | 1% |
| VH872139 | FT | 0.56 | -0.30 | 2% |
| VH872143 | FT | 0.68 | -1.03 | 2% |
| VH872146 | FT | 0.64 | -0.77 | 2% |
| VH872197 | FT | 0.88 | -2.32 | 1% |
| VH872206 | FT | 0.83 | -1.96 | 1% |
| VH872213 | FT | 0.93 | -3.01 | 1% |
| VH872216 | FT | 0.70 | -1.07 | 2% |
| VH872218 | FT | 0.45 | 0.22 | 2% |
| VH872222 | FT | 0.65 | -0.79 | 1% |
| VH872685 | FT | 0.47 | 0.15 | 2% |
| VH872690 | FT | 0.29 | 1.04 | 3% |
| VH872899 | FT | 0.67 | -0.87 | 2% |
| VH872909 | FT | 0.58 | -0.34 | 1% |
| VH872925 | FT | 0.74 | -1.28 | 2% |
| VH872937 | FT | 0.74 | -1.27 | 2% |
| VH872941 | FT | 0.80 | -1.68 | 2% |
| VH873004 | FT | 0.53 | -0.17 | 2% |
| VH873010 | FT | 0.67 | -0.97 | 2% |
| VH873930 | FT | 0.86 | -2.18 | 7% |
| VH873935 | FT | 0.80 | -1.66 | 1% |
| VH873990 | FT | 0.47 | 0.14 | 1% |
| VH873991 | FT | 0.77 | -1.51 | 2% |
| VH874021 | FT | 0.38 | 0.59 | 2% |
| VH874484 | FT | 0.74 | -1.26 | 2% |
| VH874548 | FT | 0.64 | -0.72 | 1% |
| VH874586 | FT | 0.58 | -0.39 | 2% |
| VH874606 | FT | 0.65 | -0.74 | 2% |
| VH874610 | FT | 0.67 | -0.90 | 2% |
| VH874612 | FT | 0.63 | -0.83 | 3% |
| VH874648 | FT | 0.41 | 0.37 | 3% |
| VH877748 | FT | 0.71 | -1.28 | 2% |
| VH887937 | FT | 0.92 | -2.76 | 2% |
| VH887965 | FT | 0.83 | -1.92 | 1% |
| VH906457 | FT | 0.57 | -0.38 | 2% |
| VH906746 | FT | 0.43 | 0.39 | 2% |
| VH914368 | FT | 0.78 | -1.52 | 2% |

Table 7.E.7 Item Difficulties and Omit Rate, High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Item ID | Item Type | *p*-value | IRT *b*-value | Omit Rate |
| VH861284 | FT | 0.65 | -0.75 | 1% |
| VH861314 | FT | 0.72 | -1.14 | 2% |
| VH861317 | FT | 0.51 | -0.08 | 3% |
| VH861594 | FT | 0.58 | -0.39 | 2% |
| VH861657 | FT | 0.41 | 0.45 | 5% |
| VH861665 | FT | 0.33 | 0.86 | 2% |
| VH861706 | FT | 0.82 | -1.77 | 2% |
| VH861715 | FT | 0.73 | -1.17 | 2% |
| VH861763 | FT | 0.85 | -1.99 | 2% |
| VH861767 | FT | 0.67 | -0.80 | 5% |
| VH861770 | FT | 0.70 | -1.02 | 3% |
| VH861882 | FT | 0.80 | -1.67 | 3% |
| VH861891 | FT | 0.52 | -0.08 | 2% |
| VH861895 | FT | 0.63 | -0.63 | 2% |
| VH861904 | FT | 0.74 | -1.27 | 3% |
| VH861926 | FT | 0.80 | -1.61 | 2% |
| VH861941 | FT | 0.75 | -1.36 | 2% |
| VH861966 | FT | 0.70 | -1.04 | 3% |
| VH861971 | FT | 0.66 | -0.80 | 3% |
| VH861982 | FT | 0.79 | -1.54 | 2% |
| VH862037 | FT | 0.66 | -0.80 | 2% |
| VH862061 | FT | 0.59 | -0.44 | 2% |
| VH863057 | FT | 0.79 | -1.57 | 1% |
| VH863073 | FT | 0.76 | -1.32 | 2% |
| VH863075 | FT | 0.36 | 0.65 | 3% |
| VH863079 | FT | 0.45 | 0.22 | 3% |
| VH863082 | FT | 0.44 | 0.35 | 3% |
| VH863087 | FT | 0.68 | -0.92 | 2% |
| VH863112 | FT | 0.57 | -0.45 | 3% |
| VH863116 | FT | 0.71 | -1.11 | 3% |
| VH863131 | FT | 0.55 | -0.22 | 2% |
| VH863164 | FT | 0.59 | -0.44 | 3% |
| VH863168 | FT | 0.44 | 0.30 | 3% |
| VH863171 | FT | 0.52 | -0.10 | 3% |
| VH863173 | FT | 0.66 | -0.76 | 2% |
| VH877096 | FT | 0.80 | -1.59 | 2% |
| VH898133 | FT | 0.62 | -0.79 | 2% |
| VH907952 | FT | 0.66 | -0.79 | 2% |
| VH908011 | FT | 0.50 | -0.01 | 2% |
| VH917976 | FT | 0.90 | -2.49 | 2% |
| VH918124 | FT | 0.75 | -1.31 | 1% |

### Appendix 7.F: Reliability Estimates

**Notes:**

* The reliabilities will be reported only for samples that comprise 11 or more examinees.
* In some cases in appendix 7.F, score reliabilities will not be estimable and will be presented in the tables as “N/A.”
* Results based on samples that contain 50 or fewer examinees should be interpreted with caution due to small sample sizes.

Table 7.F.1 Reliabilities and Standard Errors of Measurement (SEMs) by Gender

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Male N | Male Reliability | Male Raw Score SEM | Female N | Female Reliability | Female Raw Score SEM |
| Grade 5 Version 1 | 1,476 | 0.84 | 2.50 | 720 | 0.82 | 2.55 |
| Grade 5 Version 2 | 1,330 | 0.86 | 2.42 | 705 | 0.86 | 2.46 |
| Grade 8 Version 1 | 1,544 | 0.88 | 2.44 | 711 | 0.85 | 2.53 |
| Grade 8 Version 2 | 1,369 | 0.87 | 2.45 | 694 | 0.85 | 2.48 |
| High school Version 1 | 2,448 | 0.85 | 2.52 | 1,246 | 0.83 | 2.55 |
| High school Version 2 | 2,211 | 0.87 | 2.51 | 1,119 | 0.84 | 2.55 |
| Grade 10 Version 1 | 71 | 0.85 | 2.48 | 39 | 0.85 | 2.54 |
| Grade 10 Version 2 | 129 | 0.88 | 2.50 | 60 | 0.88 | 2.49 |
| Grade 11 Version 1 | 771 | 0.86 | 2.49 | 377 | 0.82 | 2.53 |
| Grade 11 Version 2 | 705 | 0.87 | 2.52 | 353 | 0.84 | 2.55 |
| Grade 12 Version 1 | 1,606 | 0.85 | 2.54 | 830 | 0.83 | 2.55 |
| Grade 12 Version 2 | 1,377 | 0.87 | 2.51 | 706 | 0.84 | 2.56 |

Table 7.F.2 Reliabilities and SEMs by Ethnicity

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | American Indian or Alaska Native N | American Indian or Alaska Native Reliability | American Indian or Alaska Native Raw Score SEM | Asian N | Asian Reliability | Asian Raw Score SEM | Native Hawaiian or Other Pacific Islander N | Native Hawaiian or Other Pacific Islander Reliability | Native Hawaiian or Other Pacific Islander Raw Score SEM |
| Grade 5 Version 1 | 13 | 0.87 | 2.30 | 153 | 0.84 | 2.59 | 12 | 0.66 | 2.64 |
| Grade 5 Version 2 | 11 | 0.94 | 2.21 | 171 | 0.85 | 2.51 | 10 | N/A | N/A |
| Grade 8 Version 1 | 12 | 0.88 | 2.53 | 155 | 0.84 | 2.59 | 15 | 0.87 | 2.52 |
| Grade 8 Version 2 | 13 | 0.82 | 2.31 | 149 | 0.85 | 2.57 | 9 | N/A | N/A |
| High school Version 1 | 22 | 0.85 | 2.35 | 296 | 0.80 | 2.63 | 21 | 0.72 | 2.60 |
| High school Version 2 | 28 | 0.87 | 2.30 | 269 | 0.82 | 2.62 | 19 | 0.73 | 2.56 |
| Grade 10 Version 1 | 0 | N/A | N/A | 1 | N/A | N/A | 0 | N/A | N/A |
| Grade 10 Version 2 | 3 | N/A | N/A | 9 | N/A | N/A | 0 | N/A | N/A |
| Grade 11 Version 1 | 9 | N/A | N/A | 90 | 0.81 | 2.62 | 6 | N/A | N/A |
| Grade 11 Version 2 | 5 | N/A | N/A | 65 | 0.85 | 2.58 | 4 | N/A | N/A |
| Grade 12 Version 1 | 13 | 0.85 | 2.32 | 205 | 0.80 | 2.63 | 15 | 0.78 | 2.60 |
| Grade 12 Version 2 | 20 | 0.87 | 2.24 | 195 | 0.81 | 2.63 | 15 | 0.67 | 2.53 |

Table 7.F.2 Reliabilities and SEMS by Ethnicity (Continued One)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Filipino N | Filipino Reliability | Filipino Raw Score SEM | Hispanic or Latino N | Hispanic or Latino Reliability | Hispanic or Latino Raw Score SEM | Black or African American N | Black or African American Reliability | Black or African American Raw Score SEM |
| Grade 5 Version 1 | 42 | 0.81 | 2.73 | 1,321 | 0.83 | 2.51 | 171 | 0.84 | 2.55 |
| Grade 5 Version 2 | 49 | 0.81 | 2.54 | 1,150 | 0.87 | 2.41 | 175 | 0.85 | 2.50 |
| Grade 8 Version 1 | 53 | 0.84 | 2.52 | 1,278 | 0.87 | 2.46 | 189 | 0.87 | 2.43 |
| Grade 8 Version 2 | 58 | 0.77 | 2.63 | 1,222 | 0.86 | 2.45 | 158 | 0.86 | 2.46 |
| High school Version 1 | 118 | 0.80 | 2.68 | 2,000 | 0.84 | 2.54 | 309 | 0.84 | 2.48 |
| High school Version 2 | 111 | 0.88 | 2.58 | 1,714 | 0.86 | 2.54 | 253 | 0.85 | 2.52 |
| Grade 10 Version 1 | 7 | N/A | N/A | 54 | 0.81 | 2.51 | 19 | 0.86 | 2.36 |
| Grade 10 Version 2 | 1 | N/A | N/A | 93 | 0.87 | 2.52 | 16 | 0.87 | 2.32 |
| Grade 11 Version 1 | 31 | 0.74 | 2.74 | 658 | 0.85 | 2.51 | 74 | 0.83 | 2.48 |
| Grade 11 Version 2 | 33 | 0.86 | 2.69 | 590 | 0.86 | 2.54 | 84 | 0.84 | 2.58 |
| Grade 12 Version 1 | 80 | 0.81 | 2.65 | 1,288 | 0.83 | 2.56 | 216 | 0.84 | 2.49 |
| Grade 12 Version 2 | 77 | 0.88 | 2.54 | 1,031 | 0.85 | 2.54 | 153 | 0.85 | 2.50 |

Table 7.F.2 Reliabilities and SEMS by Ethnicity (Continued Two)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | White N | White Reliability | White Raw Score SEM | Two or More Races N | Two or More Races Reliability | Two or More Races Raw Score SEM |
| Grade 5 Version 1 | 390 | 0.85 | 2.47 | 82 | 0.82 | 2.52 |
| Grade 5 Version 2 | 382 | 0.85 | 2.43 | 70 | 0.87 | 2.41 |
| Grade 8 Version 1 | 471 | 0.87 | 2.45 | 71 | 0.84 | 2.55 |
| Grade 8 Version 2 | 374 | 0.88 | 2.43 | 65 | 0.82 | 2.50 |
| High school Version 1 | 797 | 0.87 | 2.47 | 111 | 0.83 | 2.50 |
| High school Version 2 | 823 | 0.88 | 2.46 | 91 | 0.85 | 2.54 |
| Grade 10 Version 1 | 20 | 0.79 | 2.49 | 8 | N/A | N/A |
| Grade 10 Version 2 | 56 | 0.89 | 2.46 | 8 | N/A | N/A |
| Grade 11 Version 1 | 246 | 0.87 | 2.43 | 27 | 0.84 | 2.39 |
| Grade 11 Version 2 | 242 | 0.88 | 2.43 | 30 | 0.85 | 2.52 |
| Grade 12 Version 1 | 531 | 0.87 | 2.49 | 76 | 0.82 | 2.52 |
| Grade 12 Version 2 | 525 | 0.87 | 2.47 | 53 | 0.86 | 2.54 |

Table 7.F.3 Reliabilities and SEMs by English Proficiency

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | English Only N | English Only Reliability | English Only Raw Score SEM | Initial Fluent English Proficient N | Initial Fluent English Proficient Reliability | Initial Fluent English Proficient Raw Score SEM |
| Grade 5 Version 1 | 1,290 | 0.84 | 2.51 | 23 | 0.87 | 2.54 |
| Grade 5 Version 2 | 1,214 | 0.86 | 2.45 | 26 | 0.80 | 2.55 |
| Grade 8 Version 1 | 1,311 | 0.87 | 2.45 | 23 | 0.81 | 2.66 |
| Grade 8 Version 2 | 1,148 | 0.87 | 2.44 | 37 | 0.83 | 2.57 |
| High school Version 1 | 2,131 | 0.86 | 2.49 | 57 | 0.81 | 2.60 |
| High school Version 2 | 2,047 | 0.87 | 2.50 | 65 | 0.86 | 2.59 |
| Grade 10 Version 1 | 71 | 0.86 | 2.45 | 1 | N/A | N/A |
| Grade 10 Version 2 | 141 | 0.87 | 2.53 | 5 | N/A | N/A |
| Grade 11 Version 1 | 639 | 0.85 | 2.47 | 16 | 0.83 | 2.63 |
| Grade 11 Version 2 | 591 | 0.87 | 2.50 | 18 | 0.81 | 2.67 |
| Grade 12 Version 1 | 1,421 | 0.86 | 2.51 | 40 | 0.80 | 2.59 |
| Grade 12 Version 2 | 1,315 | 0.86 | 2.50 | 42 | 0.89 | 2.52 |

Table 7.F.3 Reliabilities and SEMs by English Proficiency (Continued One)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | English Learner N | English Learner Reliability | English Learner Raw Score SEM | Reclassified Fluent English Proficient N | Reclassified Fluent English Proficient Reliability | Reclassified Fluent English Proficient Raw Score SEM |
| Grade 5 Version 1 | 669 | 0.83 | 2.52 | 210 | 0.78 | 2.55 |
| Grade 5 Version 2 | 649 | 0.86 | 2.41 | 146 | 0.85 | 2.40 |
| Grade 8 Version 1 | 642 | 0.86 | 2.49 | 279 | 0.86 | 2.48 |
| Grade 8 Version 2 | 565 | 0.85 | 2.51 | 313 | 0.86 | 2.42 |
| High school Version 1 | 899 | 0.84 | 2.57 | 603 | 0.81 | 2.58 |
| High school Version 2 | 698 | 0.85 | 2.60 | 514 | 0.84 | 2.54 |
| Grade 10 Version 1 | 12 | 0.90 | 2.45 | 26 | 0.75 | 2.63 |
| Grade 10 Version 2 | 17 | 0.91 | 2.41 | 26 | 0.85 | 2.37 |
| Grade 11 Version 1 | 292 | 0.85 | 2.54 | 198 | 0.84 | 2.54 |
| Grade 11 Version 2 | 248 | 0.86 | 2.57 | 198 | 0.83 | 2.56 |
| Grade 12 Version 1 | 595 | 0.83 | 2.59 | 379 | 0.79 | 2.60 |
| Grade 12 Version 2 | 433 | 0.84 | 2.62 | 290 | 0.85 | 2.54 |

Table 7.F.3 Reliabilities and SEMs by English Proficiency (Continued Two)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | To Be Determined N | To Be Determined Reliability | To Be Determined Raw Score SEM | English Proficiency Unknown N | English Proficiency Unknown Reliability | English Proficiency Unknown Raw Score SEM |
| Grade 5 Version 1 | 0 | N/A | N/A | 4 | N/A | N/A |
| Grade 5 Version 2 | 0 | N/A | N/A | 0 | N/A | N/A |
| Grade 8 Version 1 | 0 | N/A | N/A | 0 | N/A | N/A |
| Grade 8 Version 2 | 0 | N/A | N/A | 0 | N/A | N/A |
| High school Version 1 | 2 | N/A | N/A | 2 | N/A | N/A |
| High school Version 2 | 1 | N/A | N/A | 5 | N/A | N/A |
| Grade 10 Version 1 | 0 | N/A | N/A | 0 | N/A | N/A |
| Grade 10 Version 2 | 0 | N/A | N/A | 0 | N/A | N/A |
| Grade 11 Version 1 | 2 | N/A | N/A | 1 | N/A | N/A |
| Grade 11 Version 2 | 0 | N/A | N/A | 3 | N/A | N/A |
| Grade 12 Version 1 | 0 | N/A | N/A | 1 | N/A | N/A |
| Grade 12 Version 2 | 1 | N/A | N/A | 2 | N/A | N/A |

Table 7.F.4 Reliabilities and SEMs by Economic Status

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Not Economically Disadvantaged N | Not Economically Disadvantaged Reliability | Not Economically Disadvantaged Raw Score SEM | Economically Disadvantaged N | Economically Disadvantaged Reliability | Economically Disadvantaged Raw Score SEM |
| Grade 5 Version 1 | 690 | 0.84 | 2.58 | 1,505 | 0.83 | 2.49 |
| Grade 5 Version 2 | 618 | 0.87 | 2.48 | 1,417 | 0.86 | 2.41 |
| Grade 8 Version 1 | 797 | 0.87 | 2.50 | 1,458 | 0.87 | 2.45 |
| Grade 8 Version 2 | 675 | 0.86 | 2.50 | 1,388 | 0.86 | 2.44 |
| High school Version 1 | 1,392 | 0.85 | 2.55 | 2,302 | 0.84 | 2.51 |
| High school Version 2 | 1,300 | 0.86 | 2.55 | 2,030 | 0.86 | 2.51 |
| Grade 10 Version 1 | 35 | 0.85 | 2.64 | 75 | 0.84 | 2.43 |
| Grade 10 Version 2 | 57 | 0.88 | 2.58 | 132 | 0.87 | 2.47 |
| Grade 11 Version 1 | 402 | 0.84 | 2.54 | 746 | 0.86 | 2.48 |
| Grade 11 Version 2 | 354 | 0.86 | 2.56 | 704 | 0.86 | 2.52 |
| Grade 12 Version 1 | 955 | 0.85 | 2.56 | 1,481 | 0.84 | 2.53 |
| Grade 12 Version 2 | 889 | 0.86 | 2.54 | 1,194 | 0.86 | 2.52 |

Table 7.F.5 Reliabilities and SEMs by Migrant Status

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Migrant N | Migrant Reliability | Migrant Raw Score SEM | Nonmigrant N | Nonmigrant Reliability | Nonmigrant Raw Score SEM |
| Grade 5 Version 1 | 9 | N/A | N/A | 2,187 | 0.84 | 2.52 |
| Grade 5 Version 2 | 22 | 0.78 | 2.28 | 2,013 | 0.86 | 2.43 |
| Grade 8 Version 1 | 18 | 0.81 | 2.30 | 2,237 | 0.87 | 2.47 |
| Grade 8 Version 2 | 15 | 0.86 | 2.43 | 2,048 | 0.86 | 2.46 |
| High school Version 1 | 14 | 0.80 | 2.67 | 3,680 | 0.85 | 2.53 |
| High school Version 2 | 9 | N/A | N/A | 3,321 | 0.86 | 2.53 |
| Grade 10 Version 1 | 0 | N/A | N/A | 110 | 0.85 | 2.50 |
| Grade 10 Version 2 | 0 | N/A | N/A | 189 | 0.88 | 2.50 |
| Grade 11 Version 1 | 4 | N/A | N/A | 1,144 | 0.85 | 2.50 |
| Grade 11 Version 2 | 4 | N/A | N/A | 1,054 | 0.86 | 2.53 |
| Grade 12 Version 1 | 10 | N/A | N/A | 2,426 | 0.84 | 2.54 |
| Grade 12 Version 2 | 5 | N/A | N/A | 2,078 | 0.86 | 2.53 |

Table 7.F.6 Reliabilities and SEMs by Primary Disabilities

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Intellectual Disability N | Intellectual Disability Reliability | Intellectual Disability Raw Score SEM | Hearing Impairment N | Hearing Impairment Reliability | Hearing Impairment Raw Score SEM | Speech or Language Impairment N | Speech or Language Impairment Reliability | Speech or Language Impairment Raw Score SEM |
| Grade 5 Version 1 | 782 | 0.79 | 2.59 | 25 | 0.88 | 2.30 | 67 | 0.75 | 2.29 |
| Grade 5 Version 2 | 729 | 0.82 | 2.52 | 17 | 0.73 | 2.44 | 49 | 0.77 | 2.23 |
| Grade 8 Version 1 | 939 | 0.85 | 2.53 | 31 | 0.82 | 2.35 | 43 | 0.82 | 2.30 |
| Grade 8 Version 2 | 836 | 0.85 | 2.50 | 17 | 0.81 | 2.52 | 36 | 0.79 | 2.34 |
| High school Version 1 | 1,661 | 0.80 | 2.60 | 41 | 0.76 | 2.46 | 36 | 0.79 | 2.09 |
| High school Version 2 | 1,573 | 0.84 | 2.56 | 22 | 0.86 | 2.56 | 26 | 0.84 | 2.37 |
| Grade 10 Version 1 | 40 | 0.78 | 2.65 | 0 | N/A | N/A | 3 | N/A | N/A |
| Grade 10 Version 2 | 95 | 0.87 | 2.45 | 2 | N/A | N/A | 0 | N/A | N/A |
| Grade 11 Version 1 | 450 | 0.81 | 2.59 | 24 | 0.83 | 2.40 | 17 | 0.72 | 1.99 |
| Grade 11 Version 2 | 469 | 0.84 | 2.57 | 6 | N/A | N/A | 12 | 0.83 | 2.24 |
| Grade 12 Version 1 | 1,171 | 0.80 | 2.60 | 17 | 0.61 | 2.54 | 16 | 0.83 | 2.18 |
| Grade 12 Version 2 | 1,009 | 0.84 | 2.57 | 14 | 0.89 | 2.52 | 14 | 0.85 | 2.47 |

Table 7.F.6 Reliabilities and SEMs by Primary Disabilities (Continued One)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Visual Impairment N | Visual Impairment Reliability | Visual Impairment Raw Score SEM | Emotional Disturbance N | Emotional Disturbance Reliability | Emotional Disturbance Raw Score SEM | Orthopedic Impairment N | Orthopedic Impairment Reliability | Orthopedic Impairment Raw Score SEM |
| Grade 5 Version 1 | 9 | N/A | N/A | 13 | 0.78 | 2.18 | 65 | 0.85 | 2.59 |
| Grade 5 Version 2 | 2 | N/A | N/A | 10 | N/A | N/A | 59 | 0.84 | 2.51 |
| Grade 8 Version 1 | 4 | N/A | N/A | 20 | 0.75 | 2.27 | 69 | 0.85 | 2.58 |
| Grade 8 Version 2 | 9 | N/A | N/A | 8 | N/A | N/A | 67 | 0.87 | 2.54 |
| High school Version 1 | 15 | 0.91 | 2.48 | 28 | 0.71 | 2.05 | 143 | 0.87 | 2.57 |
| High school Version 2 | 14 | 0.92 | 2.47 | 20 | 0.92 | 1.97 | 147 | 0.85 | 2.55 |
| Grade 10 Version 1 | 0 | N/A | N/A | 1 | N/A | N/A | 3 | N/A | N/A |
| Grade 10 Version 2 | 0 | N/A | N/A | 1 | N/A | N/A | 5 | N/A | N/A |
| Grade 11 Version 1 | 6 | N/A | N/A | 9 | N/A | N/A | 51 | 0.83 | 2.70 |
| Grade 11 Version 2 | 4 | N/A | N/A | 5 | N/A | N/A | 38 | 0.80 | 2.71 |
| Grade 12 Version 1 | 9 | N/A | N/A | 18 | 0.72 | 2.08 | 89 | 0.88 | 2.51 |
| Grade 12 Version 2 | 10 | N/A | N/A | 14 | 0.91 | 2.05 | 104 | 0.87 | 2.49 |

Table 7.F.6 Reliabilities and SEMs by Primary Disabilities (Continued Two)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Other Health Impairment N | Other Health Impairment Reliability | Other Health Impairment Raw Score SEM | Specific Learning Disability N | Specific Learning Disability Reliability | Specific Learning Disability Raw Score SEM | Deaf-Blindness N | Deaf-Blindness Reliability | Deaf-Blindness Raw Score SEM |
| Grade 5 Version 1 | 128 | 0.82 | 2.36 | 194 | 0.59 | 2.14 | 1 | N/A | N/A |
| Grade 5 Version 2 | 137 | 0.85 | 2.27 | 185 | 0.76 | 1.90 | 0 | N/A | N/A |
| Grade 8 Version 1 | 129 | 0.84 | 2.30 | 195 | 0.78 | 2.14 | 0 | N/A | N/A |
| Grade 8 Version 2 | 125 | 0.79 | 2.38 | 166 | 0.72 | 2.11 | 0 | N/A | N/A |
| High school Version 1 | 155 | 0.85 | 2.27 | 240 | 0.83 | 2.11 | 0 | N/A | N/A |
| High school Version 2 | 148 | 0.88 | 2.31 | 134 | 0.83 | 2.05 | 0 | N/A | N/A |
| Grade 10 Version 1 | 8 | N/A | N/A | 17 | 0.75 | 2.14 | 0 | N/A | N/A |
| Grade 10 Version 2 | 7 | N/A | N/A | 5 | N/A | N/A | 0 | N/A | N/A |
| Grade 11 Version 1 | 61 | 0.86 | 2.28 | 102 | 0.83 | 2.10 | 0 | N/A | N/A |
| Grade 11 Version 2 | 60 | 0.87 | 2.32 | 53 | 0.88 | 1.99 | 0 | N/A | N/A |
| Grade 12 Version 1 | 86 | 0.84 | 2.24 | 121 | 0.83 | 2.12 | 0 | N/A | N/A |
| Grade 12 Version 2 | 81 | 0.88 | 2.31 | 76 | 0.77 | 2.13 | 0 | N/A | N/A |

Table 7.F.6 Reliabilities and SEMs by Primary Disabilities (Continued Three)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Multiple Disabilities N | Multiple Disabilities Reliability | Multiple Disabilities Raw Score SEM | Autism N | Autism Reliability | Autism Raw Score SEM |
| Grade 5 Version 1 | 76 | 0.82 | 2.74 | 827 | 0.84 | 2.54 |
| Grade 5 Version 2 | 75 | 0.90 | 2.50 | 766 | 0.86 | 2.48 |
| Grade 8 Version 1 | 64 | 0.81 | 2.66 | 750 | 0.87 | 2.48 |
| Grade 8 Version 2 | 62 | 0.85 | 2.60 | 727 | 0.86 | 2.48 |
| High school Version 1 | 138 | 0.87 | 2.57 | 1,204 | 0.84 | 2.55 |
| High school Version 2 | 94 | 0.88 | 2.57 | 1,130 | 0.86 | 2.55 |
| Grade 10 Version 1 | 1 | N/A | N/A | 35 | 0.87 | 2.53 |
| Grade 10 Version 2 | 7 | N/A | N/A | 65 | 0.87 | 2.57 |
| Grade 11 Version 1 | 36 | 0.87 | 2.45 | 381 | 0.85 | 2.54 |
| Grade 11 Version 2 | 21 | 0.84 | 2.49 | 381 | 0.86 | 2.57 |
| Grade 12 Version 1 | 101 | 0.87 | 2.61 | 788 | 0.84 | 2.56 |
| Grade 12 Version 2 | 66 | 0.89 | 2.57 | 684 | 0.86 | 2.54 |

Table 7.F.6 Reliabilities and SEMs by Primary Disabilities (Continued Four)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | Traumatic Brain Injury N | Traumatic Brain Injury Reliability | Traumatic Brain Injury Raw Score SEM | Not Classified N | Not Classified Reliability | Not Classified Raw Score SEM |
| Grade 5 Version 1 | 9 | N/A | N/A | 0 | N/A | N/A |
| Grade 5 Version 2 | 6 | N/A | N/A | 0 | N/A | N/A |
| Grade 8 Version 1 | 11 | 0.87 | 2.12 | 0 | N/A | N/A |
| Grade 8 Version 2 | 10 | N/A | N/A | 0 | N/A | N/A |
| High school Version 1 | 33 | 0.87 | 2.23 | 0 | N/A | N/A |
| High school Version 2 | 22 | 0.85 | 2.51 | 0 | N/A | N/A |
| Grade 10 Version 1 | 2 | N/A | N/A | 0 | N/A | N/A |
| Grade 10 Version 2 | 2 | N/A | N/A | 0 | N/A | N/A |
| Grade 11 Version 1 | 11 | 0.93 | 1.99 | 0 | N/A | N/A |
| Grade 11 Version 2 | 9 | N/A | N/A | 0 | N/A | N/A |
| Grade 12 Version 1 | 20 | 0.82 | 2.36 | 0 | N/A | N/A |
| Grade 12 Version 2 | 11 | 0.83 | 2.26 | 0 | N/A | N/A |

### Appendix 7.G: Validity Analyses

**Note:** \* p < 0.01

Table 7.G.1 Correlations Between Raw Scores and Test Engagement Response

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Grade Level | PT Raw Score Mean | Level of Engagement Mean Response | SD PT Raw Score | SD Engagement Response | N | Correlation |
| Grade 5 PT 1 (Life Sciences) | 8.22 | 2.58 | 2.72 | 0.68 | 4,589 | 0.63\* |
| Grade 5 PT 2 (Physical Sciences) | 8.37 | 2.58 | 2.82 | 0.67 | 4,516 | 0.63\* |
| Grade 5 PT 3 (Earth and Space Sciences) | 6.82 | 2.53 | 2.51 | 0.69 | 4,505 | 0.52\* |
| Grade 8 PT 1 (Life Sciences) | 7.96 | 2.59 | 2.64 | 0.67 | 4,704 | 0.61\* |
| Grade 8 PT 2 (Physical Sciences) | 7.20 | 2.58 | 2.72 | 0.68 | 4,614 | 0.59\* |
| Grade 8 PT 3 (Earth and Space Sciences) | 7.36 | 2.57 | 2.90 | 0.69 | 4,635 | 0.63\* |
| High school PT 1 (Life Sciences) | 7.59 | 2.66 | 2.79 | 0.64 | 7,484 | 0.63\* |
| High school PT 2 (Physical Sciences) | 7.77 | 2.64 | 2.85 | 0.65 | 7,469 | 0.65\* |
| High school PT 3 (Earth and Space Sciences) | 6.85 | 2.62 | 2.64 | 0.67 | 7,476 | 0.57\* |
| Grade 10 PT 1 (Life Sciences) | 7.90 | 2.64 | 2.69 | 0.67 | 321 | 0.53\* |
| Grade 10 PT 2 (Physical Sciences) | 7.84 | 2.61 | 3.07 | 0.69 | 330 | 0.58\* |
| Grade 10 PT 3 (Earth and Space Sciences) | 6.97 | 2.61 | 2.61 | 0.68 | 321 | 0.52\* |
| Grade 11 PT 1 (Life Sciences) | 7.70 | 2.66 | 2.73 | 0.63 | 2,324 | 0.61\* |
| Grade 11 PT 2 (Physical Sciences) | 7.88 | 2.61 | 2.81 | 0.66 | 2,323 | 0.62\* |
| Grade 11 PT 3 (Earth and Space Sciences) | 7.01 | 2.62 | 2.58 | 0.65 | 2,315 | 0.51\* |
| Grade 12 PT 1 (Life Sciences) | 7.51 | 2.66 | 2.83 | 0.65 | 4,839 | 0.66\* |
| Grade 12 PT 2 (Physical Sciences) | 7.72 | 2.65 | 2.86 | 0.65 | 4,816 | 0.67\* |
| Grade 12 PT 3 (Earth and Space Sciences) | 6.77 | 2.62 | 2.66 | 0.68 | 4,840 | 0.59\* |

Table 7.G.2 Raw Score by PT Engagement Response, Grade Five

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Performance Task | Level of Engagement | Mean Raw Score | SD Raw Score | N |
| PT 1 (Life Sciences) Version 1 | Fully Engaged | 8.96 | 2.07 | 1,663 |
| PT 1 (Life Sciences) Version 1 | Moderately Engaged | 7.15 | 2.39 | 472 |
| PT 1 (Life Sciences) Version 1 | Minimally Engaged | 4.12 | 2.72 | 231 |
| PT 1 (Life Sciences) Version 2 | Fully Engaged | 9.37 | 2.06 | 1,499 |
| PT 1 (Life Sciences) Version 2 | Moderately Engaged | 7.16 | 2.29 | 462 |
| PT 1 (Life Sciences) Version 2 | Minimally Engaged | 4.44 | 2.83 | 261 |
| PT 2 (Physical Sciences) Version 1 | Fully Engaged | 9.45 | 2.21 | 1,599 |
| PT 2 (Physical Sciences) Version 1 | Moderately Engaged | 7.21 | 2.48 | 512 |
| PT 2 (Physical Sciences) Version 1 | Minimally Engaged | 4.57 | 2.76 | 224 |
| PT 2 (Physical Sciences) Version 2 | Fully Engaged | 9.25 | 2.32 | 1,459 |
| PT 2 (Physical Sciences) Version 2 | Moderately Engaged | 7.06 | 2.27 | 485 |
| PT 2 (Physical Sciences) Version 2 | Minimally Engaged | 4.38 | 2.40 | 236 |
| PT 3 (Earth and Space Sciences) Version 1 | Fully Engaged | 7.46 | 2.21 | 1,513 |
| PT 3 (Earth and Space Sciences) Version 1 | Moderately Engaged | 6.07 | 2.13 | 560 |
| PT 3 (Earth and Space Sciences) Version 1 | Minimally Engaged | 4.47 | 2.53 | 258 |
| PT 3 (Earth and Space Sciences) Version 2 | Fully Engaged | 7.63 | 2.27 | 1,397 |
| PT 3 (Earth and Space Sciences) Version 2 | Moderately Engaged | 6.18 | 2.12 | 512 |
| PT 3 (Earth and Space Sciences) Version 2 | Minimally Engaged | 4.00 | 2.50 | 264 |

Table 7.G.3 Raw Score by PT Engagement Response, Grade Eight

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Performance Task | Level of Engagement | Mean Raw Score | SD Raw Score | N |
| PT 1 (Life Sciences) Version 1 | Fully Engaged | 8.72 | 2.13 | 1,731 |
| PT 1 (Life Sciences) Version 1 | Moderately Engaged | 6.95 | 2.05 | 487 |
| PT 1 (Life Sciences) Version 1 | Minimally Engaged | 4.74 | 2.87 | 231 |
| PT 1 (Life Sciences) Version 2 | Fully Engaged | 8.83 | 2.25 | 1,557 |
| PT 1 (Life Sciences) Version 2 | Moderately Engaged | 6.87 | 2.07 | 430 |
| PT 1 (Life Sciences) Version 2 | Minimally Engaged | 4.35 | 2.82 | 267 |
| PT 2 (Physical Sciences) Version 1 | Fully Engaged | 8.27 | 2.44 | 1,690 |
| PT 2 (Physical Sciences) Version 1 | Moderately Engaged | 6.16 | 2.31 | 476 |
| PT 2 (Physical Sciences) Version 1 | Minimally Engaged | 4.26 | 2.52 | 244 |
| PT 2 (Physical Sciences) Version 2 | Fully Engaged | 7.78 | 2.25 | 1,490 |
| PT 2 (Physical Sciences) Version 2 | Moderately Engaged | 5.99 | 2.19 | 442 |
| PT 2 (Physical Sciences) Version 2 | Minimally Engaged | 3.88 | 2.54 | 271 |
| PT 3 (Earth and Space Sciences) Version 1 | Fully Engaged | 8.25 | 2.62 | 1,697 |
| PT 3 (Earth and Space Sciences) Version 1 | Moderately Engaged | 5.69 | 2.55 | 472 |
| PT 3 (Earth and Space Sciences) Version 1 | Minimally Engaged | 3.84 | 2.16 | 250 |
| PT 3 (Earth and Space Sciences) Version 2 | Fully Engaged | 8.43 | 2.35 | 1,501 |
| PT 3 (Earth and Space Sciences) Version 2 | Moderately Engaged | 6.20 | 2.17 | 424 |
| PT 3 (Earth and Space Sciences) Version 2 | Minimally Engaged | 4.06 | 2.36 | 291 |

Table 7.G.4 Raw Score by PT Engagement Response, High School

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Performance Task | Level of Engagement | Mean Raw Score | SD Raw Score | N |
| PT 1 (Life Sciences) Version 1 | Fully Engaged | 8.22 | 2.28 | 2,949 |
| PT 1 (Life Sciences) Version 1 | Moderately Engaged | 6.29 | 2.28 | 637 |
| PT 1 (Life Sciences) Version 1 | Minimally Engaged | 3.68 | 2.78 | 348 |
| PT 1 (Life Sciences) Version 2 | Fully Engaged | 8.53 | 2.33 | 2,689 |
| PT 1 (Life Sciences) Version 2 | Moderately Engaged | 6.08 | 2.30 | 514 |
| PT 1 (Life Sciences) Version 2 | Minimally Engaged | 3.44 | 2.80 | 347 |
| PT 2 (Physical Sciences) Version 1 | Fully Engaged | 8.47 | 2.23 | 2,870 |
| PT 2 (Physical Sciences) Version 1 | Moderately Engaged | 6.62 | 2.23 | 689 |
| PT 2 (Physical Sciences) Version 1 | Minimally Engaged | 3.57 | 2.80 | 371 |
| PT 2 (Physical Sciences) Version 2 | Fully Engaged | 8.80 | 2.41 | 2,597 |
| PT 2 (Physical Sciences) Version 2 | Moderately Engaged | 6.47 | 2.28 | 592 |
| PT 2 (Physical Sciences) Version 2 | Minimally Engaged | 3.38 | 2.72 | 350 |
| PT 3 (Earth and Space Sciences) Version 1 | Fully Engaged | 7.62 | 2.31 | 2,841 |
| PT 3 (Earth and Space Sciences) Version 1 | Moderately Engaged | 6.06 | 2.23 | 678 |
| PT 3 (Earth and Space Sciences) Version 1 | Minimally Engaged | 3.90 | 2.73 | 413 |
| PT 3 (Earth and Space Sciences) Version 2 | Fully Engaged | 7.41 | 2.32 | 2,575 |
| PT 3 (Earth and Space Sciences) Version 2 | Moderately Engaged | 5.86 | 2.02 | 594 |
| PT 3 (Earth and Space Sciences) Version 2 | Minimally Engaged | 3.41 | 2.65 | 375 |

Table 7.G.5 Raw Score by PT Engagement Response, Grade Ten

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Performance Task | Level of Engagement | Mean Raw Score | SD Raw Score | N |
| PT 1 (Life Sciences) Version 1 | Fully Engaged | 8.45 | 2.24 | 91 |
| PT 1 (Life Sciences) Version 1 | Moderately Engaged | 7.29 | 2.28 | 17 |
| PT 1 (Life Sciences) Version 1 | Minimally Engaged | 5.79 | 3.70 | 14 |
| PT 1 (Life Sciences) Version 2 | Fully Engaged | 8.60 | 2.35 | 148 |
| PT 1 (Life Sciences) Version 2 | Moderately Engaged | 6.57 | 1.85 | 30 |
| PT 1 (Life Sciences) Version 2 | Minimally Engaged | 4.38 | 3.23 | 21 |
| PT 2 (Physical Sciences) Version 1 | Fully Engaged | 8.63 | 2.15 | 89 |
| PT 2 (Physical Sciences) Version 1 | Moderately Engaged | 7.41 | 2.44 | 22 |
| PT 2 (Physical Sciences) Version 1 | Minimally Engaged | 4.20 | 4.28 | 15 |
| PT 2 (Physical Sciences) Version 2 | Fully Engaged | 8.66 | 2.68 | 151 |
| PT 2 (Physical Sciences) Version 2 | Moderately Engaged | 7.14 | 2.40 | 29 |
| PT 2 (Physical Sciences) Version 2 | Minimally Engaged | 3.21 | 2.86 | 24 |
| PT 3 (Earth and Space Sciences) Version 1 | Fully Engaged | 7.68 | 2.30 | 90 |
| PT 3 (Earth and Space Sciences) Version 1 | Moderately Engaged | 6.44 | 3.08 | 16 |
| PT 3 (Earth and Space Sciences) Version 1 | Minimally Engaged | 5.18 | 3.05 | 17 |
| PT 3 (Earth and Space Sciences) Version 2 | Fully Engaged | 7.58 | 2.28 | 142 |
| PT 3 (Earth and Space Sciences) Version 2 | Moderately Engaged | 5.68 | 1.97 | 38 |
| PT 3 (Earth and Space Sciences) Version 2 | Minimally Engaged | 3.44 | 2.48 | 18 |

Table 7.G.6 Raw Score by PT Engagement Response, Grade Eleven

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Performance Task | Level of Engagement | Mean Raw Score | SD Raw Score | N |
| PT 1 (Life Sciences) Version 1 | Fully Engaged | 8.37 | 2.27 | 907 |
| PT 1 (Life Sciences) Version 1 | Moderately Engaged | 6.48 | 2.38 | 207 |
| PT 1 (Life Sciences) Version 1 | Minimally Engaged | 4.11 | 2.73 | 101 |
| PT 1 (Life Sciences) Version 2 | Fully Engaged | 8.54 | 2.35 | 823 |
| PT 1 (Life Sciences) Version 2 | Moderately Engaged | 6.02 | 2.14 | 193 |
| PT 1 (Life Sciences) Version 2 | Minimally Engaged | 3.76 | 2.76 | 93 |
| PT 2 (Physical Sciences) Version 1 | Fully Engaged | 8.62 | 2.26 | 879 |
| PT 2 (Physical Sciences) Version 1 | Moderately Engaged | 6.75 | 2.21 | 216 |
| PT 2 (Physical Sciences) Version 1 | Minimally Engaged | 4.15 | 2.85 | 123 |
| PT 2 (Physical Sciences) Version 2 | Fully Engaged | 8.85 | 2.41 | 778 |
| PT 2 (Physical Sciences) Version 2 | Moderately Engaged | 6.59 | 2.27 | 221 |
| PT 2 (Physical Sciences) Version 2 | Minimally Engaged | 3.86 | 2.79 | 106 |
| PT 3 (Earth and Space Sciences) Version 1 | Fully Engaged | 7.85 | 2.26 | 865 |
| PT 3 (Earth and Space Sciences) Version 1 | Moderately Engaged | 6.25 | 2.30 | 217 |
| PT 3 (Earth and Space Sciences) Version 1 | Minimally Engaged | 4.32 | 2.75 | 123 |
| PT 3 (Earth and Space Sciences) Version 2 | Fully Engaged | 7.36 | 2.37 | 797 |
| PT 3 (Earth and Space Sciences) Version 2 | Moderately Engaged | 6.07 | 2.08 | 215 |
| PT 3 (Earth and Space Sciences) Version 2 | Minimally Engaged | 3.82 | 2.53 | 98 |

Table 7.G.7 Raw Score by PT Engagement Response, Grade Twelve

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Performance Task | Level of Engagement | Mean Raw Score | SD Raw Score | N |
| PT 1 (Life Sciences) Version 1 | Fully Engaged | 8.14 | 2.28 | 1,951 |
| PT 1 (Life Sciences) Version 1 | Moderately Engaged | 6.15 | 2.21 | 413 |
| PT 1 (Life Sciences) Version 1 | Minimally Engaged | 3.36 | 2.67 | 233 |
| PT 1 (Life Sciences) Version 2 | Fully Engaged | 8.52 | 2.32 | 1,718 |
| PT 1 (Life Sciences) Version 2 | Moderately Engaged | 6.08 | 2.44 | 291 |
| PT 1 (Life Sciences) Version 2 | Minimally Engaged | 3.22 | 2.77 | 233 |
| PT 2 (Physical Sciences) Version 1 | Fully Engaged | 8.40 | 2.22 | 1,902 |
| PT 2 (Physical Sciences) Version 1 | Moderately Engaged | 6.52 | 2.22 | 451 |
| PT 2 (Physical Sciences) Version 1 | Minimally Engaged | 3.22 | 2.62 | 233 |
| PT 2 (Physical Sciences) Version 2 | Fully Engaged | 8.78 | 2.39 | 1,668 |
| PT 2 (Physical Sciences) Version 2 | Moderately Engaged | 6.34 | 2.27 | 342 |
| PT 2 (Physical Sciences) Version 2 | Minimally Engaged | 3.16 | 2.65 | 220 |
| PT 3 (Earth and Space Sciences) Version 1 | Fully Engaged | 7.51 | 2.32 | 1,886 |
| PT 3 (Earth and Space Sciences) Version 1 | Moderately Engaged | 5.95 | 2.15 | 445 |
| PT 3 (Earth and Space Sciences) Version 1 | Minimally Engaged | 3.64 | 2.66 | 273 |
| PT 3 (Earth and Space Sciences) Version 2 | Fully Engaged | 7.42 | 2.30 | 1,636 |
| PT 3 (Earth and Space Sciences) Version 2 | Moderately Engaged | 5.76 | 1.98 | 341 |
| PT 3 (Earth and Space Sciences) Version 2 | Minimally Engaged | 3.25 | 2.71 | 259 |

Table 7.G.8 Total Testing Time (In Minutes) by Grade and Version

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Grade Level and Version | Number | Mean | Standard Deviation | Minimum | Maximum | Percentile Points 1 | Percentile Points 10 | Percentile Points 25 | Percentile Points 50 | Percentile Points 75 | Percentile Points 90 | Percentile Points 99 |
| Grade 5 Version 1 | 2,654 | 11.34 | 7.48 | 0.10 | 75.12 | 0.26 | 4.01 | 6.59 | 9.93 | 14.51 | 20.09 | 36.46 |
| Grade 5 Version 2 | 2,476 | 12.29 | 8.66 | 0.10 | 103.75 | 0.30 | 3.90 | 6.97 | 10.66 | 15.48 | 22.30 | 41.69 |
| Grade 8 Version 1 | 2,682 | 13.84 | 9.63 | 0.12 | 111.87 | 0.24 | 4.34 | 7.91 | 12.03 | 17.65 | 24.71 | 46.75 |
| Grade 8 Version 2 | 2,534 | 12.95 | 8.64 | 0.08 | 79.20 | 0.23 | 3.65 | 7.39 | 11.36 | 16.64 | 24.06 | 41.24 |
| High school Version 1 | 4,706 | 12.98 | 9.05 | 0.10 | 116.71 | 0.16 | 1.57 | 7.83 | 11.90 | 16.87 | 23.14 | 43.85 |
| High school Version 2 | 4,108 | 13.83 | 9.13 | 0.10 | 85.15 | 0.19 | 2.31 | 8.31 | 12.60 | 18.09 | 25.02 | 43.70 |
| Grade 10 Version 1 | 143 | 13.79 | 10.69 | 0.13 | 79.49 | 0.13 | 2.43 | 8.38 | 11.86 | 16.68 | 24.44 | 60.96 |
| Grade 10 Version 2 | 234 | 13.39 | 8.42 | 0.17 | 50.27 | 0.22 | 2.11 | 8.21 | 12.79 | 17.51 | 24.98 | 38.56 |
| Grade 11 Version 1 | 1,396 | 12.84 | 8.47 | 0.10 | 116.71 | 0.16 | 2.84 | 8.19 | 11.68 | 16.25 | 22.19 | 39.98 |
| Grade 11 Version 2 | 1,264 | 13.30 | 8.69 | 0.14 | 69.86 | 0.24 | 2.64 | 8.02 | 11.85 | 17.39 | 23.84 | 38.85 |
| Grade 12 Version 1 | 3,167 | 13.00 | 9.22 | 0.10 | 82.55 | 0.15 | 1.32 | 7.61 | 12.04 | 17.08 | 23.34 | 44.72 |
| Grade 12 Version 2 | 2,610 | 14.12 | 9.38 | 0.10 | 85.15 | 0.18 | 2.16 | 8.49 | 12.86 | 18.56 | 25.55 | 44.13 |

Table 7.G.9 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Five

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Student Performance Quartile | Raw Score Range | Number of Students | Mean Testing Time | Standard Deviation | Minimum | Maximum | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Version 1 Q 1 | 0–15 | 617 | 7.53 | 7.93 | 0.10 | 56.97 | 0.14 | 0.67 | 1.98 | 5.25 | 10.51 | 16.70 | 36.46 |
| Version 1 Q 2 | 16–22 | 636 | 12.57 | 8.08 | 1.94 | 75.12 | 3.45 | 5.52 | 7.04 | 10.41 | 15.75 | 21.71 | 37.70 |
| Version 1 Q 3 | 23–27 | 651 | 12.11 | 6.13 | 2.88 | 42.16 | 4.27 | 6.20 | 7.51 | 10.40 | 15.09 | 20.35 | 32.36 |
| Version 1 Q 4 | 28–36 | 750 | 12.75 | 6.53 | 3.66 | 55.81 | 4.60 | 6.35 | 8.38 | 11.29 | 15.38 | 20.33 | 37.44 |
| Version 2 Q 1 | 0–15 | 576 | 8.52 | 10.01 | 0.10 | 86.85 | 0.17 | 0.69 | 1.91 | 5.64 | 11.66 | 18.75 | 52.63 |
| Version 2 Q 2 | 16–22 | 643 | 13.53 | 8.83 | 2.20 | 103.75 | 3.05 | 5.77 | 8.14 | 11.17 | 16.48 | 23.55 | 38.32 |
| Version 2 Q 3 | 23–27 | 532 | 13.46 | 7.79 | 1.88 | 93.13 | 3.96 | 6.21 | 8.47 | 11.61 | 16.39 | 22.67 | 38.30 |
| Version 2 Q 4 | 28–36 | 725 | 13.34 | 6.98 | 1.93 | 52.72 | 3.88 | 6.41 | 8.54 | 11.75 | 16.23 | 22.29 | 39.38 |

Table 7.G.10 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Eight

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Student Performance Quartile | Raw Score Range | Number of Students | Mean Testing Time | Standard Deviation | Minimum | Maximum | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Version 1 Q 1 | 0–14 | 574 | 9.36 | 10.49 | 0.12 | 104.19 | 0.15 | 0.52 | 1.95 | 6.47 | 12.87 | 21.63 | 45.41 |
| Version 1 Q 2 | 15–21 | 765 | 14.25 | 9.01 | 1.63 | 85.41 | 2.80 | 5.73 | 8.12 | 12.17 | 17.77 | 25.13 | 46.75 |
| Version 1 Q 3 | 22–27 | 575 | 15.53 | 9.62 | 1.61 | 111.87 | 3.33 | 7.17 | 9.34 | 13.03 | 19.12 | 26.87 | 46.32 |
| Version 1 Q 4 | 28–36 | 768 | 15.50 | 8.49 | 2.67 | 78.24 | 4.23 | 7.62 | 10.04 | 13.78 | 18.24 | 24.69 | 45.58 |
| Version 2 Q 1 | 0–14 | 602 | 7.96 | 8.01 | 0.08 | 57.87 | 0.10 | 0.48 | 1.71 | 5.93 | 11.41 | 17.75 | 34.98 |
| Version 2 Q 2 | 15–20 | 620 | 13.92 | 8.93 | 1.48 | 79.20 | 2.59 | 5.76 | 7.86 | 11.67 | 17.14 | 25.11 | 45.09 |
| Version 2 Q 3 | 21–27 | 661 | 14.45 | 7.73 | 1.83 | 61.88 | 3.52 | 6.85 | 8.96 | 12.68 | 17.28 | 24.28 | 41.05 |
| Version 2 Q 4 | 28–36 | 651 | 15.10 | 8.01 | 1.68 | 64.03 | 4.81 | 7.39 | 9.53 | 12.96 | 18.12 | 25.51 | 40.99 |

Table 7.G.11 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, High School

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Student Performance Quartile | Raw Score Range | Number of Students | Mean Testing Time | Standard Deviation | Minimum | Maximum | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Version 1 Q 1 | 0–13 | 1,136 | 5.97 | 9.01 | 0.10 | 82.55 | 0.13 | 0.19 | 0.62 | 2.66 | 7.83 | 15.87 | 50.90 |
| Version 1 Q 2 | 14–20 | 1,215 | 14.41 | 7.84 | 0.91 | 62.45 | 2.13 | 7.06 | 9.33 | 12.47 | 17.78 | 24.30 | 42.96 |
| Version 1 Q 3 | 21–25 | 1,017 | 15.94 | 8.03 | 1.34 | 71.62 | 4.71 | 8.47 | 10.61 | 14.04 | 19.38 | 25.98 | 44.57 |
| Version 1 Q 4 | 26–36 | 1,338 | 15.37 | 7.66 | 1.31 | 116.71 | 5.40 | 8.49 | 10.65 | 13.55 | 18.09 | 23.51 | 41.08 |
| Version 2 Q 1 | 0–13 | 925 | 6.37 | 7.98 | 0.10 | 85.15 | 0.14 | 0.30 | 0.93 | 3.58 | 9.15 | 16.00 | 36.45 |
| Version 2 Q 2 | 14–20 | 1,109 | 15.32 | 8.53 | 0.91 | 73.62 | 1.69 | 6.86 | 9.66 | 13.70 | 19.06 | 25.99 | 43.43 |
| Version 2 Q 3 | 21–26 | 927 | 17.02 | 8.65 | 2.42 | 69.86 | 5.09 | 8.76 | 11.04 | 14.88 | 20.96 | 27.50 | 50.50 |
| Version 2 Q 4 | 27–36 | 1,147 | 15.81 | 7.56 | 2.11 | 63.49 | 4.65 | 8.68 | 10.83 | 13.93 | 18.93 | 25.07 | 42.40 |

Table 7.G.12 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Ten

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Student Performance Quartile | Raw Score Range | Number of Students | Mean Testing Time | Standard Deviation | Minimum | Maximum | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Version 1 Q 1 | 0–15 | 34 | 8.62 | 15.18 | 0.13 | 79.49 | 0.13 | 0.17 | 1.67 | 2.86 | 9.06 | 20.63 | 79.49 |
| Version 1 Q 2 | 16–22 | 37 | 16.20 | 10.40 | 7.36 | 60.96 | 7.36 | 8.38 | 9.82 | 12.25 | 17.24 | 26.08 | 60.96 |
| Version 1 Q 3 | 23–27 | 35 | 15.87 | 7.78 | 5.03 | 39.21 | 5.03 | 9.08 | 10.50 | 12.93 | 20.89 | 25.42 | 39.21 |
| Version 1 Q 4 | 28–36 | 37 | 14.18 | 6.22 | 6.72 | 36.52 | 6.72 | 7.64 | 10.58 | 12.68 | 16.23 | 21.49 | 36.52 |
| Version 2 Q 1 | 0–13 | 57 | 5.61 | 6.56 | 0.17 | 35.10 | 0.17 | 0.49 | 0.93 | 3.11 | 8.64 | 12.89 | 23.28 |
| Version 2 Q 2 | 14–19 | 52 | 15.35 | 8.76 | 2.94 | 50.27 | 2.94 | 6.82 | 9.76 | 12.45 | 18.61 | 23.25 | 43.43 |
| Version 2 Q 3 | 20–26 | 55 | 17.15 | 7.76 | 3.73 | 38.56 | 3.73 | 7.97 | 12.10 | 14.64 | 22.25 | 28.15 | 34.14 |
| Version 2 Q 4 | 27–36 | 70 | 15.31 | 5.70 | 4.80 | 37.06 | 4.80 | 9.91 | 11.33 | 14.12 | 18.00 | 23.76 | 27.12 |

Table 7.G.13 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Eleven

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Student Performance Quartile | Raw Score Range | Number of Students | Mean Testing Time | Standard Deviation | Minimum | Maximum | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Version 1 Q 1 | 0–14 | 313 | 6.63 | 7.45 | 0.10 | 37.97 | 0.11 | 0.22 | 0.97 | 4.02 | 9.51 | 17.17 | 32.73 |
| Version 1 Q 2 | 15–20 | 331 | 13.48 | 7.39 | 0.91 | 52.31 | 1.84 | 6.95 | 8.71 | 11.44 | 16.03 | 22.77 | 38.46 |
| Version 1 Q 3 | 21–26 | 342 | 14.97 | 7.47 | 3.65 | 71.62 | 5.97 | 8.31 | 10.13 | 12.99 | 17.66 | 23.22 | 44.07 |
| Version 1 Q 4 | 27–36 | 410 | 15.27 | 8.49 | 3.87 | 116.71 | 4.96 | 8.21 | 10.46 | 13.51 | 17.66 | 23.19 | 41.08 |
| Version 2 Q 1 | 0–14 | 300 | 6.80 | 7.48 | 0.14 | 47.60 | 0.16 | 0.36 | 1.21 | 4.55 | 9.80 | 16.91 | 36.28 |
| Version 2 Q 2 | 15–20 | 313 | 14.41 | 7.60 | 1.65 | 55.98 | 1.69 | 6.06 | 9.03 | 12.88 | 18.58 | 24.54 | 33.71 |
| Version 2 Q 3 | 21–26 | 289 | 16.77 | 9.38 | 3.36 | 69.86 | 4.59 | 8.33 | 10.55 | 14.27 | 20.82 | 27.20 | 54.83 |
| Version 2 Q 4 | 27–36 | 362 | 14.94 | 7.01 | 3.51 | 45.70 | 4.84 | 8.12 | 10.13 | 13.33 | 17.39 | 23.52 | 38.53 |

Table 7.G.14 Distribution of Total Testing Time (In Minutes) at Each Quartile Group by Version, Grade Twelve

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Student Performance Quartile | Raw Score Range | Number of Students | Mean Testing Time | Standard Deviation | Minimum | Maximum | 1st Percentile | 10th Percentile | 25th Percentile | 50th Percentile | 75th Percentile | 90th Percentile | 99th Percentile |
| Version 1 Q 1 | 0–12 | 785 | 5.41 | 8.89 | 0.10 | 82.55 | 0.13 | 0.18 | 0.52 | 2.08 | 6.83 | 14.28 | 51.49 |
| Version 1 Q 2 | 13–19 | 727 | 14.79 | 8.18 | 1.17 | 62.45 | 1.80 | 6.95 | 9.57 | 12.91 | 18.09 | 24.50 | 44.72 |
| Version 1 Q 3 | 20–25 | 840 | 16.04 | 7.97 | 1.34 | 63.04 | 3.93 | 8.52 | 10.73 | 14.04 | 19.38 | 26.44 | 44.57 |
| Version 1 Q 4 | 26–36 | 815 | 15.59 | 7.41 | 1.31 | 64.25 | 5.22 | 8.72 | 10.83 | 13.71 | 18.47 | 24.08 | 41.79 |
| Version 2 Q 1 | 0–13 | 609 | 6.65 | 8.43 | 0.10 | 85.15 | 0.13 | 0.27 | 0.92 | 3.80 | 9.60 | 16.37 | 41.28 |
| Version 2 Q 2 | 14–20 | 693 | 15.86 | 8.94 | 0.91 | 73.62 | 1.64 | 7.17 | 9.93 | 14.06 | 19.36 | 26.81 | 44.06 |
| Version 2 Q 3 | 21–26 | 593 | 17.13 | 8.34 | 2.42 | 61.18 | 5.48 | 8.97 | 11.34 | 15.19 | 20.96 | 27.53 | 50.15 |
| Version 2 Q 4 | 27–36 | 715 | 16.31 | 7.95 | 2.11 | 63.49 | 4.32 | 8.93 | 10.94 | 14.35 | 19.65 | 25.86 | 42.52 |

## Surveys

Three separate surveys were developed and administered to collect additional information on the California Alternate Assessment (CAA) for Science field testing in 2018–2019:

1. A student engagement survey completed by test examiners about each student’s level of engagement with each embedded performance task (PT)
2. A student survey answered by students about their reactions to each embedded PT
3. A test examiner survey completed by test examiners to solicit their feedback on the CAA for Science administration overall.

### Survey Design and Questionnaire Development

The three surveys were designed and developed by members of the Educational Testing Service validity research team with extensive experience in designing and developing student and teacher surveys.

#### Student Survey

Student survey responses, which were provided by the test examiner, were collected from local educational agencies via the California Assessment of Student Performance and Progress (CAASPP) test delivery system for every embedded PT administered to every student. After an embedded PT was administered to the student, the student’s test examiner would then respond to a survey consisting of three questions.

1. How engaged the student was with the embedded PT just administered?
2. Were Orienting Activity #1 and the first five questions individualized?
3. Were Orienting Activity #2 and the last five questions individualized?

Table 8.A.1 presents the responses for how engaged the student was with the embedded PT. Most students were fully engaged with the embedded PT just administered, while some students were either moderately engaged or minimally engaged with the embedded PT.

Table 8.A.2 and table 8.A.3 provide the number of students whose administration was individualized by the test examiner. The majority of students did not receive an individualization.

#### Test Examiner Survey

To gain insights from the field for potential future improvement, an optional survey was presented to test examiners to obtain teachers’ feedback on the field test administration and assessment processes overall. This survey was linked on the CAASPP website and hosted on SurveyGizmo.com, a website with survey-creation and hosting services.

### Test Examiner Survey Results

Up to 22 questions were asked on the test examiner survey. The first question asked whether or not the respondent was a test examiner for the 2018–2019 CAA for Science. Of the 248 responses to the first question, 196 indicated “Yes.” The other 52 responses indicated that the respondent had not been a CAA for Science test examiner, which ended the survey.

In the 2018–2019 administration year, there were roughly 3,000 active test examiners, indicating a 12 percent response rate from all available test examiners to the test examiner survey. Because of the small number of test examiners responding to the survey questions, the information in table 8.1 through table 8.6 have limited generalizability; caution should be used when interpreting the test examiner survey results.

About 15 percent of respondents chose to individualize the test, and these individualizations were important to their student’s performance on the assessment. Generally, test examiners found the instructions to be clear or reasonably clear.

Table 8.1 through table 8.6 provide the results for the test examiner surveys over all grade levels.

#### Responses to Background Questions

Table 8.1 displays responses to the question, “Are you the primary teacher for the students you tested this year?” The data shows that just over 67 percent of test examiners who responded were also the teacher for all of the students they tested; therefore, most test examiners are likely to be the student’s teacher and familiar with the needs of the student.

Table 8.1 Teacher for Student Tested

|  |  |  |
| --- | --- | --- |
| Are you the primary teacher for the students you tested this year? | Test Examiners | Percent |
| 1. Yes, I am the teacher for all of the students I tested this year. | 125 | 67% |
| 1. Yes, I am the teacher for some of the students I tested this year. | 34 | 18% |
| 1. No, I am not the teacher for any of the students I tested this year. | 27 | 15% |

Table 8.2 displays responses to the question, “How many of your students attempted the 2018–19 CAA for Science field test?” The data shows that about half of the test examiners who responded administered the test to one to five students. Nearly a third of respondents indicated that they administered the test to 16 or more students.

Table 8.2 How Many Students Tested Per Test Examiner

|  |  |  |
| --- | --- | --- |
| How many of your students attempted the 2018–‍19 CAA for Science field test? | Test Examiners | Percent |
| 1. 1–5 students | 90 | 48% |
| 1. 6–10 students | 26 | 14% |
| 1. 11–15 students | 11 | 6% |
| 1. 16 or more students | 60 | 32% |

Table 8.3 displays the distribution of grades for which test examiners who responded administered the CAA for Science in response to the question, “In what grade(s) did you administer the PTs? (Select all that apply.)”

Table 8.3 Grade Administered

|  |  |  |
| --- | --- | --- |
| In what grade(s) did you administer the PTs? (Select all that apply.) | Test Examiners | Percent |
| Grade five | 93 | 49% |
| Grade eight | 62 | 33% |
| High school | 70 | 37% |

#### Responses Regarding Individualization

Test examiners were asked about how frequently they individualized the test for their students. Roughly 85 percent indicated no individualization. Therefore, the test examiners who responded were not likely to use the option to individualize, which is reflected in the low percentages of students receiving an individualization (refer to table 5.1 through table 5.3). All responses are displayed in table 8.4.

Table 8.4 Frequency of Individualization

|  |  |  |
| --- | --- | --- |
| Did you take advantage of the option to “individualize” a performance task for any of your students? | Test Examiners | Percent |
| 1. Yes, I individualized a performance task. | 25 | 15% |
| 1. No, I did not individualize a performance task. | 137 | 85% |

#### Responses Regarding Helpfulness of the *Administration Planning Guide*

Approximately 70 percent of respondents indicated that the *Administration Planning Guide* was either very or somewhat helpful. Thus, the *Administration Planning Guide* was helpful to the test examiners who used the guide. Very few test examiners who responded did not find the *Administration Planning Guide* helpful. However, over one quarter of the test examiners who responded either did not know about or did not use the *Administration Planning Guide* although it was available. The results are displayed in table 8.5.

Table 8.5 Helpfulness of the *Administration Planning Guide*

|  |  |  |
| --- | --- | --- |
| How helpful was the *Administration Planning Guide* in preparing you to administer the test? | Test Examiners | Percent |
| Very helpful | 51 | 36% |
| Somewhat helpful | 47 | 33% |
| Not helpful | 7 | 5% |
| Did not access | 19 | 13% |
| Did not know it was available | 18 | 13% |

#### Responses Regarding the Move to Online Testing

Approximately 71 percent of respondents indicated that moving to online testing was preferable. All test examiners who responded to the survey and who answered this question confirmed that they had delivered the test previously in paper form. After experiencing both the paper and online formats, test examiners preferred the online format and felt that it was an improvement from the paper format. The results are displayed in table 8.6.

Table 8.6 Preference for Online Format

|  |  |  |
| --- | --- | --- |
| Was the new online format of the of the performance task within the test delivery system an improvement over the old method of administration? | Test Examiners | Percent |
| Yes | 55 | 71% |
| No | 8 | 8% |
| Unsure | 14 | 14% |

### Appendix 8.A: Distribution of Student Survey Responses

#### Survey for Responsive Students

Table 8.A.1 Responses to Question 1, “How engaged was the student with this performance task?”

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Embedded Performance Task | Number of Responses | Fully Engaged | Moderately Engaged | Minimally Engaged |
| Grade 5, PT 1: Life Sciences | 4686 | 68% | 20% | 11% |
| Grade 5, PT 2: Physical Sciences | 4548 | 68% | 22% | 10% |
| Grade 5, PT 3: Earth and Space Sciences | 4516 | 65% | 24% | 12% |
| Grade 8, PT 1: Life Sciences | 4776 | 70% | 20% | 11% |
| Grade 8, PT 2: Physical Sciences | 4642 | 69% | 20% | 11% |
| Grade 8, PT 3: Earth and Space Sciences | 4655 | 69% | 19% | 12% |
| High school, PT 1: Life Sciences | 7542 | 75% | 15% | 9% |
| High school, PT 2: Physical Sciences | 7619 | 73% | 17% | 10% |
| High school, PT 3: Earth and Space Sciences | 7500 | 72% | 17% | 11% |

Table 8.A.2 Responses to Question 2, “Did you individualize any aspect of Orienting Activity #1 and the first five test questions, where permitted?”

|  |  |  |  |
| --- | --- | --- | --- |
| Embedded Performance Task | Number of Responses | No | Yes |
| Grade 5, PT 1: Life Sciences | 4,643 | 92% | 8% |
| Grade 5, PT 2: Physical Sciences | 4,536 | 93% | 7% |
| Grade 5, PT 3: Earth and Space Sciences | 4,492 | 93% | 7% |
| Grade 8, PT 1: Life Sciences | 4,752 | 91% | 9% |
| Grade 8, PT 2: Physical Sciences | 4,600 | 90% | 10% |
| Grade 8, PT 3: Earth and Space Sciences | 4,628 | 91% | 9% |
| High school, PT 1: Life Sciences | 7,484 | 90% | 10% |
| High school, PT 2: Physical Sciences | 7,590 | 86% | 14% |
| High school, PT 3: Earth and Space Sciences | 7,442 | 90% | 10% |

Table 8.A.3 Responses to Question 3, “Did you individualize any aspect of Orienting Activity #2 and the first five test questions, where permitted?”

|  |  |  |  |
| --- | --- | --- | --- |
| Embedded Performance Task | Number of Responses | No | Yes |
| Grade 5, PT 1: Life Sciences | 4,596 | 93% | 7% |
| Grade 5, PT 2: Physical Sciences | 4,468 | 92% | 8% |
| Grade 5, PT 3: Earth and Space Sciences | 4,428 | 93% | 7% |
| Grade 8, PT 1: Life Sciences | 4,696 | 93% | 7% |
| Grade 8, PT 2: Physical Sciences | 4,537 | 92% | 8% |
| Grade 8, PT 3: Earth and Space Sciences | 4,574 | 93% | 7% |
| High school, PT 1: Life Sciences | 7,437 | 91% | 9% |
| High school, PT 2: Physical Sciences | 7,516 | 91% | 9% |
| High school, PT 3: Earth and Space Sciences | 7,368 | 91% | 9% |

## Embedded Performance Task and Test Comparability Considerations

This chapter describes the analysis conducted to evaluate the impact of both the choice of materials and individualization of the assessment on the embedded performance tasks (PTs) administered as part of the 2018–2019 California Alternate Assessment (CAA) for Science field test. The results of this analysis are summarized in this chapter. Also refer to subsection [*5.5 Accessibility Features for the Field Test*](#_Accessibility_Features_for), where individualizations are further described.

### Considerations

The CAA for Science field test, which occurred during the 2018–2019 California Assessment of Student Performance and Progress (CAASPP) administration, provided Educational Testing Service (ETS) an opportunity to collect data about this test to inform psychometric decisions.

One aspect of the CAA for Science field test design was the flexibility offered to test examiners to choose the type of materials used in conducting the science activities. The rationale for providing choice to test examiners was to enable them to create testing conditions that were representative of classroom instruction. However, there were concerns about the potential impact of giving test examiners the flexibility to choose materials to conduct activities associated with the embedded PTs. Specifically, will this choice result in differential performance on the items associated with the California Next Generation Science Standards (CA NGSS) Core Content Connectors (Science Connectors)? To answer this question, ETS conducted an evaluation to determine the impact on student performance by the decision to individualize and by the choice of materials used by the test examiner.

The test examiners downloaded the secure embedded PTs *Directions for Administration (DFAs)*, which contained descriptions of the hands-on activity for each Science Connector within the embedded PT (California Department of Education, 2019). The *DFAs* included the directions for the hands-on activities, associated test questions, and recommended materials for each exemplar activity. Additionally, if individualizations or flexibility toward the exemplar activity were acceptable, then suggestions for alternative materials were provided.

During the test administration, the test examiner entered the student response into the CAASPP test delivery system and responded to the survey questions to note the use of alternative materials and scripts for each Science Connector or activity and the amount of student engagement—fully engaged, moderately engaged, or minimally engaged.

For the material choices analysis, the ETS Psychometric Analysis & Research team used the statistical analysis sample, as described in subsection [*7.1.2 Sample for the Analyses*](#_Sample_for_the)*,* to examine the relationship between the choice of materials used and Science Connector scores. The dependent variable for each of the models was the Science Connector score, which is the sum of all points earned on items associated with each Science Connector. The Science Connector scores were analyzed because the questions regarding individualizations were asked at the Science Connector level rather than the embedded PT level.

The independent variables of interest are as follows:

1. **Disability type—**This variable indicates the specific student disability and was collected during test registration.
2. **Use of individualization—**This variable is a test examiner–reported measure indicating whether certain aspects of the Science Connector or activity (e.g., directions) were individualized to make the content accessible to the student. Its measurement uses a coded variable (yes or no).
3. **Student engagement—**This variable is a test examiner–reported measure, completed after the administration of each embedded PT, where the test examiner indicated whether the student was highly, moderately, or minimally engaged.
4. **Choice of materials—**This variable is a test examiner–reported measure indicating whether alternative materials, in lieu of the exemplar materials, were used for a particular Science Connector. Its measurement uses a coded variable (yes or no).

These investigations were conducted only on groups of at least 25 students. For example, in grade five, only 12 students had a visual impairment; these students were excluded from the analysis.

After reviewing the number of students who received an alternative material, several linear models were estimated for each embedded PT to evaluate the incremental impact of the test examiners’ material choices on the Science Connector scores. The linear models were only estimated when there were at least 100 students in the statistical sample who received any type of alternative material choice. The first four models evaluate the main effects of the independent variables. In the fifth model, the interactions were computed only for models where there was a significant main effect corresponding to the material choice.

The linear models were estimated using the SAS® PROC GLM module (general linear model). The SAS® PROC GLM module will convert the categorial independent variables by creating the binary dummy variables as part of the regression computation. [[7]](#footnote-8)

Table 9.1 summarizes the sequential models that were estimated. The main effects and interactions were evaluated using the type III sum of squares (SS3). The SS3 is calculated with respect to all the other variables included in the model; the resulting sum of squares for each variable is its effect after all other variables have been accounted for. The result for each variable is equivalent to what is obtained using the type I sum of squares, when that variable is the last variable entered in the model.

Table 9.1 Linear Models Estimated in Material Choices Analysis

|  |  |  |
| --- | --- | --- |
| Model | Dependent Variable | Independent Variables |
| 1 | Science Connector score | * Disability |
| 2 | Science Connector score | * Disability * Individualization |
| 3 | Science Connector score | * Disability * Individualization * Engagement |
| 4 | Science Connector score | * Disability * Individualization * Engagement * Materials |
| 5 | Science Connector score | * Disability * Individualization * Engagement * Materials * Interaction |

The disability types were recoded into groups based on the nature of the disabilities; the disability groups are listed in table 9.2. The regression analyses were initially run with the disability types as reported during the test registration process. The regression analyses were conducted for Intellectual Disabilities versus Autism and for Intellectual Disabilities versus Learning Disabilities.

Table 9.2 Regroupings of the Disability Types

|  |  |  |
| --- | --- | --- |
| Disability Group | Reference Group | Focal Group |
| **Intellectual Disability Versus Autism** | Intellectual Disability group includes:   * Intellectual disability * Multiple disabilities * Traumatic brain injury | * Autism |
| **Intellectual Disability Versus Learning Disability** | Intellectual Disability group includes:   * Intellectual disability * Multiple disabilities * Traumatic brain injury | Learning Disability group includes:   * Emotional disturbance * Orthopedic impairment * Other health impairment * Specific learning disability * Speech or language impairment |

For each of the estimated models, the coefficient of determination, or R-squared (R2), was calculated. In addition, the change in R2 (R2 increment) was calculated to compare the differences in the increasingly complex statistical models. The R2 increment provides a convenient way to summarize the additional proportion of variance in Science Connector scores explained by adding each independent variable into the model.

### Summary of Findings from the Choice of Materials and Level of Individualization Analysis

#### Individualization Analysis

The number and percentage of students in the statistical analysis sample receiving an individualization is provided in [appendix 9.A](#_Appendix_9.A:_Choice). Very few students—1 percent or less—included in the analyses for this investigation received an individualized script or an alternative graphic (picture or diagram). Few students—5 percent or less—received alternative materials to use during the administration of the embedded PTs.

#### Model Analysis

The linear models were run to analyze the relationship between the choice of materials and the students’ Science Connector scores. Summaries of the analyzed models are provided in [appendix 9.B](#_Appendix_9.B:_Model), [appendix 9.C](#_Appendix_9.C:_Model) (for Intellectual Disabilities versus Autism), and [appendix 9.D](#_Appendix_9.D:_Model) (for Intellectual Disabilities versus Learning Disabilities). The information included in these tables are as follows:

* Model: The variables included in the model
* R2: The proportion of the variance of the Science Connector score explained by the independent variables specified in the model
* Difference in R2: The difference in R2 between the current model and the previously run model
* Significance tests: The F ratio and *p*-value testing the null hypothesis that adding an individual variable to the model does not increase the proportion of the variance explained by the model
* Partial eta-square (η2): The proportion of variance accounted for by adding a variable to the model. This can be evaluated using the following rules (Cohen, 1988):

Small effect: 0.01

Medium effect: 0.06

Large effect: 0.14

For some Science Connectors, fewer than 100 students were in the statistical sample that received an alternative material choice; for these Science Connectors, no linear models were estimated.

#### Results of the Individualization and Model Analyses

This subsection summarizes the results presented in [appendix 9.B](#_Appendix_9.B:_Model), starting with the value of the R2 and the differences in R2 across the models.

Overall, the proportion of variance in the students’ Science Connector scores accounted for by the final models was low to moderate, with the R2 ranging from 0.0705 to 0.3017. The value of R2 increased minimally by including material choice in the model; the increase in R2 ranged from 0.0000 to 0.0006. Therefore, adding material choice had negligible impact on the models.

In the final models conducted for each Science Connector, material choice accounted for a small amount of the variance. In the final models, the partial η2 for material choice ranged between 0.0000 and 0.0008. The overall magnitude of R2 increment might be small due to the very low utilization rates of material choices.

Adding student engagement to the models increased the value of R2 and the amount of variance accounted for in the students’ Science Connector scores; the increase in R2 ranged between 0.0691 and 0.2377. Additionally, for the final model run for each Science Connector, student engagement accounted for more variance in the Science Connector scores than any other variable included in the model. In the final models conducted for each Science Connector, the partial η2 for student engagement ranged from 0.0696 to 0.2381. Partial η2 values of 0.14 or greater (using the rules from Cohen, 1988) indicate that the student engagement has a large effect on the student’s scores.

##### Grade Five Model Analysis

For grade five, only the second Physical Sciences Connector had more than 100 students receiving an alternative material choice. The results of the models conducted are provided in table 9.B.1, table 9.C.1, and table 9.D.1.

The final model performed for this Science Connector included disability, individualization (e.g., alternative script), student engagement, and material choice. Therefore, the choice to use individualized materials was not significant and did not explain a significant proportion of the variance in the students’ Science Connector scores, given the other variables included in the model (partial η2 for materials ranged from 0.0000 to 0.0002).

##### Grade Eight Model Analysis

For grade eight, only the Physical Sciences Connectors had more than 100 students receiving an alternative material. The results of the models conducted are provided in table 9.B.2 and table 9.B.3, table 9.C.2 and table 9.C.3 for the Intellectual Disabilities versus Autism groups; and table 9.D.2 and table 9.D.3 for the Intellectual Disabilities versus Learning Disabilities groups.

The final models performed for these two Science Connectors included disability, individualization (e.g., alternative script), student engagement, and material choice. Therefore, the choice to use individualized materials was not significant and did not explain a significant proportion of the variance in the students’ Science Connector scores, given the other variables included in the models (partial η2 for materials ranged from 0.0000 and 0.0008).

##### High School Model Analysis

For high school, four Science Connectors had more than 100 students receiving an alternative material choice, with the first Physical Sciences Connector having more than 300 students receiving an alternative material. The four embedded PTs were the first Physical Sciences Connector, the second Physical Sciences Connector, the first Life Sciences Connector, and the first Earth and Space Sciences Connector.

The results of the linear models are presented in table 9.B.4, table 9.B.5, table 9.B.6, and table 9.B.7 when the primary disability was not grouped; table 9.C.4, table 9.C.5, table 9.C.6, and table 9.C.7 for Intellectual Disabilities versus Autism; and table 9.D.4, table 9.D.5, table 9.D.6, and table 9.D.7 for Intellectual Disabilities versus Learning Disabilities. The choice of using individualized materials did not explain a significant proportion of the variance in the students’ Science Connector scores (the partial η2 for materials ranged from 0.0000 and 0.0006).

### Implications for Future Test Administrations

#### Key Findings

There are several key findings from the evaluation of the material choices.

First, in general, test examiners did not use individualizations when administering the field test embedded PTs, particularly the options of individualizing the script or providing alternative diagrams or pictures. Additionally, few students received the use of alternative materials, usually less than 3 percent; 5 percent of the students received alternative materials for the first Physical Sciences Connector for high school.

Second, in general, individualizations and material choice do not explain a significant proportion of the variance of the students’ Science Connector scores, given the other variables in the model. For all the analyses, student engagement and disability explained significant proportions of the Science Connector scores, given the other variables in the model. Student engagement explained more variance in the Science Connector scores, which is reflected by the partial η2; student engagement had larger partial η2 values than disability, material choice, and individualizations.

Third, the amount of variance explained by the material choice was small, which is reflected in the small increases in R2 by including material choice in the model and in the small values of partial η2.

When interpreting the results of the material choice analyses, caution should be taken due to the small percentage of students who received an individualization (e.g., individualized script) or who received individualized materials. Due to the low rates of students receiving an individualization, there is low statistical power to detect possible effects of the choice to use individualized materials if an effect exists. Additionally, the test examiner chose to use individualizations or individualized materials to make the Science Connector orienting activity more accessible to the student and was based on the needs of the student. Therefore, the results of these analyses are nested within student disability and the needs of the student.

#### Recommendations

For the 2019–2020 CAA for Science operational administration, the ETS psychometric team recommends that information on individualizations—use of individualized scripts, diagrams, pictures, and materials—be collected from the test examiner. Although the number of students receiving an alternative material was low for the field test administration, as test examiners become more familiar with the format of the CAA for Science, the use of individualizations may increase.

The ETS team also recommends that science content experts who are familiar with the CAA for Science student population review the operational tasks to determine what types of individualizations might be the most appropriate for the student population. However, these recommended individualizations should not alter the content or the underlying CA NGSS Science Connector being assessed by the test questions.

### References

California Department of Education. (2019). *California Alternate Assessment for Science directions for administration, training performance task, fossils and chemical changes.* Sacramento, CA: California Department of Education.

Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

SAS Institute Inc. (2015). The GLM procedure. *SAS/STAT(4) 14.1 User’s Guide.* Cary, NC: SAS Institute Inc.

### Appendix 9.A: Choice of Materials and Individualization Analysis Data

Table 9.A.1 Individualizations—Grade Five

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Individualization | PT 1 (Life Sciences), Activity 1—Number | PT 1 (Life Sciences), Activity 1—Percent of Total | PT 1 (Life Sciences), Activity 2—Number | PT 1 (Life Sciences), Activity 2—Percent of Total | PT 2 (Physical Sciences), Activity 1—Number | PT 2 (Physical Sciences), Activity 1—Percent of Total | PT 2 (Physical Sciences), Activity 2—Number | PT 2 (Physical Sciences), Activity 2—Percent of Total | PT 3 (Earth and Space Sciences), Activity 1—Number | PT 3 (Earth and Space Sciences), Activity 1—Percent of Total | PT 3 (Earth and Space Sciences), Activity 2—Number | PT 3 (Earth and Space Sciences), Activity 2—Percent of Total |
| Using Standardized Scripts | 4,392 | 100% | 4,395 | 100% | 4,387 | 99% | 4,391 | 99% | 4,389 | 99% | 4,388 | 99% |
| Using Individualized Scripts | 22 | 0% | 19 | 0% | 27 | 1% | 23 | 1% | 25 | 1% | 26 | 1% |
| Using Standardized Graphic | 4,390 | 99% | 4,372 | 99% | 4,392 | 100% | 4,407 | 100% | 4,394 | 100% | 4,403 | 100% |
| Using Individualized Graphic | 24 | 1% | 42 | 1% | 22 | 0% | 7 | 0% | 20 | 0% | 11 | 0% |
| Using Standardized Materials | 4,352 | 99% | 4,387 | 99% | 4,383 | 99% | 4,277 | 97% | 4,401 | 100% | 4,342 | 98% |
| Using Individualized Materials | 62 | 1% | 27 | 1% | 31 | 1% | 137 | 3% | 13 | 0% | 72 | 2% |

Table 9.A.2 Individualizations—Grade Eight

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Individualization | PT 1 (Life Sciences), Activity 1—Number | PT 1 (Life Sciences), Activity 1—Percent of Total | PT 1 (Life Sciences), Activity 2—Number | PT 1 (Life Sciences), Activity 2—Percent of Total | PT 2 (Physical Sciences), Activity 1—Number | PT 2 (Physical Sciences), Activity 1—Percent of Total | PT 2 (Physical Sciences), Activity 2—Number | PT 2 (Physical Sciences), Activity 2—Percent of Total | PT 3 (Earth and Space Sciences), Activity 1—Number | PT 3 (Earth and Space Sciences), Activity 1—Percent of Total | PT 3 (Earth and Space Sciences), Activity 2—Number | PT 3 (Earth and Space Sciences), Activity 2—Percent of Total |
| Using Standardized Scripts | 4,501 | 100% | 4,500 | 100% | 4,493 | 99% | 4,494 | 99% | 4,502 | 100% | 4,500 | 100% |
| Using Individualized Scripts | 19 | 0% | 20 | 0% | 27 | 1% | 26 | 1% | 18 | 0% | 20 | 0% |
| Using Standardized Graphic | 4,489 | 99% | 4,499 | 100% | 4,505 | 100% | 4,513 | 100% | 4,508 | 100% | 4,513 | 100% |
| Using Individualized Graphic | 31 | 1% | 21 | 0% | 15 | 0% | 7 | 0% | 12 | 0% | 7 | 0% |
| Using Standardized Materials | 4,492 | 99% | 4,501 | 100% | 4,374 | 97% | 4,380 | 97% | 4,441 | 98% | 4,510 | 100% |
| Using Individualized Materials | 28 | 1% | 19 | 0% | 146 | 3% | 140 | 3% | 79 | 2% | 10 | 0% |

Table 9.A.3 Individualizations—High School

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Individualization | PT 1 (Life Sciences), Activity 1—Number | PT 1 (Life Sciences), Activity 1—Percent of Total | PT 1 (Life Sciences), Activity 2—Number | PT 1 (Life Sciences), Activity 2—Percent of Total | PT 2 (Physical Sciences), Activity 1—Number | PT 2 (Physical Sciences), Activity 1—Percent of Total | PT 2 (Physical Sciences), Activity 2—Number | PT 2 (Physical Sciences), Activity 2—Percent of Total | PT 3 (Earth and Space Sciences), Activity 1—Number | PT 3 (Earth and Space Sciences), Activity 1—Percent of Total | PT 3 (Earth and Space Sciences), Activity 2—Number | PT 3 (Earth and Space Sciences), Activity 2—Percent of Total |
| Using Standardized Scripts | 7,237 | 100% | 7,236 | 100% | 7,234 | 100% | 7,235 | 100% | 7,236 | 100% | 7,235 | 100% |
| Using Individualized Scripts | 27 | 0% | 28 | 0% | 30 | 0% | 29 | 0% | 28 | 0% | 29 | 0% |
| Using Standardized Graphic | 7,251 | 100% | 7,245 | 100% | 7,261 | 100% | 7,262 | 100% | 7,250 | 100% | 7,248 | 100% |
| Using Individualized Graphic | 13 | 0% | 19 | 0% | 3 | 0% | 2 | 0% | 14 | 0% | 16 | 0% |
| Using Standardized Materials | 7,145 | 98% | 7,236 | 100% | 6,894 | 95% | 7,110 | 98% | 7,136 | 98% | 7,206 | 99% |
| Using Individualized Materials | 119 | 2% | 28 | 0% | 370 | 5% | 154 | 2% | 128 | 2% | 58 | 1% |

### Appendix 9.B: Model Analysis Summaries

Table 9.B.1 Model Summary—Grade Five, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0955 | N/A | Disability: F, 7 = 65.53 (<0.0001) | Disability: 0.0955 |
| 2 | Disability + Individualization | 0.0957 | 0.0002 | Disability: F, 7 = 65.53 (<0.0001)  Script: F, 1 = 1.11 (0.2928) | Disability: 0.0955  Script: 0.0003 |
| 3 | Disability + Individualization + Engagement | 0.3016 | 0.2059 | Disability: F, 7 = 52.46 (<0.0001)  Script: F, 1 = 0.05 (0.8234)  Engagement: F, 2 = 616.37 (<0.0001) | Disability: 0.0786  Script: 0.0000  Engagement: 0.2226 |
| 4 | Disability + Individualization + Engagement + Materials | 0.3017 | 0.0001 | Disability: F, 7 = 52.17 (<0.0001)  Script: F, 1 = 0.08 (0.7819)  Engagement: F, 2 = 614.79 (<0.0001)  Material: F, 1 = 0.87 (0.3497) | Disability: 0.0782  Script: 0.0000  Engagement: 0.2222  Material: 0.0002 |

Table 9.B.2 Model Summary—Grade Eight, Physical Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0742 | N/A | Disability: F,8 = 44.62 (<0.0001) | Disability: 0.0742 |
| 2 | Disability + Individualization | 0.0742 | 0.0000 | Disability: F,8 = 44.60 (<0.0001)  Script: F,1 = 0.02 (0.8759) | Disability: 0.0742  Script: 0.0000 |
| 3 | Disability + Individualization + Engagement | 0.2196 | 0.1454 | Disability: F,8 = 28.88 (<0.0001)  Script: F,1 = 0.82 (0.3652)  Engagement: F,2 = 406.46 (<0.0001) | Disability: 0.0499  Script: 0.0002  Engagement: 0.1559 |
| 4 | Disability + Individualization + Engagement + Materials | 0.2199 | 0.0003 | Disability: F,8 = 29.02 (<0.0001)  Script: F,1 = 0.79 (0.3736)  Engagement: F,2 = 407.29 (<0.0001)  Material: F,1 = 1.58 (0.2092) | Disability: 0.0501  Script: 0.0002  Engagement: 0.1562  Material: 0.0004 |

Table 9.B.3 Model Summary—Grade Eight, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0721 | N/A | Disability: F, 8 = 43.23 (<0.0001) | Disability: 0.0721 |
| 2 | Disability + Individualization | 0.0721 | 0.0000 | Disability: F, 8 = 43.16 (<0.0001)  Script: F, 1 = 0.01 (0.9172) | Disability: 0.0720  Script: 0.0000 |
| 3 | Disability + Individualization + Engagement | 0.1672 | 0.0951 | Disability: F, 8 = 31.74 (<0.0001)  Script: F, 1 = 0.10 (0.7520)  Engagement: F,2 = 251.00 (<0.0001) | Disability: 0.0545  Script: 0.0000  Engagement: 0.1024 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1672 | 0.0000 | Disability: F, 8 = 31.71 (<0.0001)  Script: F, 1 = 0.10 (0.7570)  Engagement: F,2 = 250.53 (<0.0001)  Material: F, 1 = 0.04 (0.8471) | Disability: 0.0545  Script: 0.0000  Engagement: 0.1022  Material: 0.0000 |

Table 9.B.4 Model Summary—High School, Life Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0484 | N/A | Disability: F, 10 = 36.57 (<0.0001) | Disability: 0.0484 |
| 2 | Disability + Individualization | 0.0485 | 0.0001 | Disability: F, 10 = 36.51 (<0.0001)  Script: F, 1 = 1.17 (0.2800) | Disability: 0.0483  Script: 0.0002 |
| 3 | Disability + Individualization + Engagement | 0.1176 | 0.0691 | Disability: F, 10 = 30.09 (<0.0001)  Script: F, 1 = 0.04 (0.8431)  Engagement: F, 2 = 278.93 (<0.0001) | Disability: 0.0407  Script: 0.0000  Engagement: 0.0728 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1178 | 0.0002 | Disability: F, 10 = 30.22 (<0.0001)  Script: F, 1 = 0.00 (0.9716)  Engagement: F, 2 = 279.98 (<0.0001)  Material: F, 1 = 2.22 (0.1364) | Disability: 0.0408  Script: 0.0000  Engagement: 0.0731  Material: 0.0003 |

Table 9.B.5 Model Summary—High School, Physical Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0380 | N/A | Disability: F,10 = 28.42 (<0.0001) | Disability: 0.0380 |
| 2 | Disability + Individualization | 0.0381 | 0.0001 | Disability: F,10 = 28.34 (<0.0001)  Script: F,1 = 1.49 (0.2229) | Disability: 0.0379  Script: 0.0002 |
| 3 | Disability + Individualization + Engagement | 0.1749 | 0.1368 | Disability: F,10 = 21.50 (<0.0001)  Script: F,1 = 0.00 (0.9495)  Engagement: F,2 = 591.44 (<0.0001) | Disability: 0.0294  Script: 0.0000  Engagement: 0.1429 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1750 | 0.0001 | Disability: F,10 = 21.56 (<0.0001)  Script: F,1 = 0.01 (0.9309)  Engagement: F,2 = 591.97 (<0.0001)  Material: F,1 = 1.19 (0.2758) | Disability: 0.0295  Script: 0.0000  Engagement: 0.1430  Material: 0.0002 |

Table 9.B.6 Model Summary—High School, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0498 | N/A | Disability: F,10 = 37.68 (<0.0001) | Disability: 0.0498 |
| 2 | Disability + Individualization | 0.0507 | 0.0009 | Disability: F,10 = 37.58 (<0.0001)  Script: F,1 = 6.76 (0.0093) | Disability: 0.0497  Script: 0.0009 |
| 3 | Disability + Individualization + Engagement | 0.2037 | 0.1530 | Disability: F,10 = 27.27 (<0.0001)  Script: F,1 = 1.10 (0.2942)  Engagement: F,2 = 687.28 (<0.0001) | Disability: 0.0370  Script: 0.0002  Engagement: 0.1623 |
| 4 | Disability + Individualization + Engagement + Materials | 0.2038 | 0.0001 | Disability: F, 10 = 27.31 (<0.0001)  Script: F, 1 = 1.10 (0.2938)  Engagement: F, 2 = 686.16 (<0.0001)  Material: F, 1 = 0.62 (0.4293) | Disability: 0.0371  Script: 0.0002  Engagement: 0.1621  Material: 0.0001 |

Table 9.B.7 Model Summary—High School, Earth and Space Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0362 | N/A | Disability: F, 10 = 27.03 (<0.0001) | Disability: 0.0362 |
| 2 | Disability + Individualization | 0.0378 | 0.0016 | Disability: F, 10 = 26.89 (<0.0001)  Script: F, 1 = 12.31 (0.0005) | Disability: 0.0360  Script: 0.0017 |
| 3 | Disability + Individualization + Engagement | 0.1332 | 0.0954 | Disability: F, 10 = 21.17 (<0.0001)  Script: F,1 = 6.64 (0.0100)  Engagement: F, 2 = 389.01 (<0.0001) | Disability: 0.0289  Script: 0.0009  Engagement: 0.0987 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1332 | 0.0000 | Disability: F, 10 = 21.16(<0.0001)  Script: F, 1 = 6.64 (0.0100)  Engagement: F, 2 = 388.85 (<0.0001)  Material: F, 1 = 0.00 (0.9734) | Disability: 0.0289  Script: 0.0009  Engagement: 0.0986  Material: 0.0000 |

### Appendix 9.C: Model Analysis Summaries for Intellectual Disabilities Versus Autism

Table 9.C.1 Model Summary—Grade Five, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0016 | N/A | Disability: F,1 = 5.39 (0.0203) | Disability: 0.0016 |
| 2 | Disability + Individualization | 0.0019 | 0.0003 | Disability: F,1 = 5.33 (0.0211)  Script: F,1 = 0.99 (0.3198) | Disability: 0.0016  Script: 0.0003 |
| 3 | Disability + Individualization + Engagement | 0.2396 | 0.2377 | Disability: F,1 = 46.23 (<0.0001)  Script: F,1 = 0.00 (0.9444)  Engagement: F,2 = 528.29 (<0.0001) | Disability: 0.0135  Script: 0.0000  Engagement: 0.2381 |
| 4 | Disability + Individualization + Engagement + Materials | 0.2396 | 0.0000 | Disability: F,1 = 46.37 (<0.0001)  Script: F,1 = 0.00 (0.9685)  Engagement: F,2 = 528.20 (<0.0001)  Material: F,1 = 0.15 (0.6966) | Disability: 0.0135  Script: 0.0000  Engagement: 0.2381  Material: 0.0000 |

Table 9.C.2 Model Summary—Grade Eight, Physical Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0000 | N/A | Disability: F,1 = 0.14 (0.7074) | Disability: 0.0000 |
| 2 | Disability + Individualization | 0.0001 | 0.0001 | Disability: F,1 = 0.13 (0.7155)  Script: F,1 = 0.10 (0.7493) | Disability: 0.0000  Script: 0.0000 |
| 3 | Disability + Individualization + Engagement | 0.1645 | 0.1644 | Disability: F,1 = 4.83 (0.0280)  Script: F,1 = 1.45 (0.2288)  Engagement: F,2 = 346.46 (<0.0001) | Disability: 0.0014  Script: 0.0004  Engagement: 0.1644 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1651 | 0.0006 | Disability: F,1 = 4.91 (0.0267)  Script: F,1 =1.39 (0.2383)  Engagement: F,2 = 347.99 (<0.0001)  Material: F,1 = 2.72 (0.0989) | Disability: 0.0014  Script: 0.0004  Engagement: 0.1651  Material: 0.0008 |

Table 9.C.3 Model Summary—Grade Eight, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0020 | N/A | Disability: F,1 = 7.27 (0.0070) | Disability: 0.0020 |
| 2 | Disability + Individualization | 0.0020 | 0.0000 | Disability: F,1 = 7.27 (0.0071)  Script: F,1 = 0.00 (0.9498) | Disability: 0.0020  Script: 0.0000 |
| 3 | Disability + Individualization + Engagement | 0.1173 | 0.1153 | Disability: F,1 = 24.29 (<0.0001)  Script: F,1 = 0.34 (0.5582)  Engagement: F,2 = 229.90 (<0.0001) | Disability: 0.0069  Script: 0.0001  Engagement: 0.1155 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1173 | 0.0000 | Disability: F,1 = 24.27 (<0.0001)  Script: F,1 = 0.35 (0.5557)  Engagement: F,2 = 228.93 (<0.0001)  Material: F,1 = 0.05 (0.8280) | Disability: 0.0068  Script: 0.0001  Engagement: 0.1151  Material: 0.0000 |

Table 9.C.4 Model Summary—High School, Life Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0007 | N/A | Disability: F,1 = 4.03 (0.0448) | Disability: 0.0007 |
| 2 | Disability + Individualization | 0.0009 | 0.0002 | Disability: F,1 = 3.95 (0.0469)  Script: F,1 = 1.13 (0.2881) | Disability: 0.0007  Script: 0.0002 |
| 3 | Disability + Individualization + Engagement | 0.0700 | 0.0691 | Disability: F,1 = 17.39 (<0.0001)  Script: F,1 = 0.01 (0.9162)  Engagement: F,2 = 221.81 (<0.0001) | Disability: 0.0029  Script: 0.0000  Engagement: 0.0692 |
| 4 | Disability + Individualization + Engagement + Materials | 0.0705 | 0.0005 | Disability: F,1 = 17.62 (<0.0001)  Script: F,1 = 0.01 (0.9304)  Engagement: F,2 = 223.14 (<0.0001)  Material: F,1 = 3.35 (0.0672) | Disability: 0.0029  Script: 0.0000  Engagement: 0.0696  Material: 0.0006 |

Table 9.C.5 Model Summary—High School, Physical Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0001 | N/A | Disability: F,1 = 0.74 (0.3892) | Disability: 0.0001 |
| 2 | Disability + Individualization | 0.0003 | 0.0002 | Disability: F,1 = 0.70 (0.4026)  Script: F,1 = 1.13 (0.2873) | Disability: 0.0001  Script: 0.0002 |
| 3 | Disability + Individualization + Engagement | 0.1421 | 0.1418 | Disability: F,1 = 18.28 (<0.0001)  Script: F,1 = 0.01 (0.9188)  Engagement: F,2 = 493.91 (<0.0001) | Disability: 0.0030  Script: 0.0000  Engagement: 0.1418 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1424 | 0.0003 | Disability: F,1 = 18.38 (<0.0001)  Script: F,1 = 0.02 (0.9020)  Engagement: F,2 = 494.83 (<0.0001)  Material: F,1 = 1.71 (0.1909) | Disability: 0.0031  Script: 0.0000  Engagement: 0.1421  Material: 0.0003 |

Table 9.C.6 Model Summary—High School, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0001 | N/A | Disability: F,1 = 0.78 (0.3778) | Disability: 0.0001 |
| 2 | Disability + Individualization | 0.0010 | 0.0009 | Disability: F,1 = 0.91 (0.3392)  Script: F,1 = 5.08 (0.0242) | Disability: 0.0002  Script: 0.0008 |
| 3 | Disability + Individualization + Engagement | 0.1632 | 0.1622 | Disability: F,1 = 6.96 (0.0084)  Script: F,1 = 0.70 (0.4032)  Engagement: F,2 = 578.87 (<0.0001) | Disability: 0.0012  Script: 0.0001  Engagement: 0.1623 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1635 | 0.0003 | Disability: F,1 = 7.08 (0.0078)  Script: F,1 = 0.71 (0.4009)  Engagement: F,2 = 579.49 (<0.0001)  Material: F,1 = 1.82 (0.1779) | Disability: 0.0012  Script: 0.0001  Engagement: 0.1624  Material: 0.0003 |

Table 9.C.7 Model Summary—High School, Earth and Space Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0005 | N/A | Disability: F,1 = 2.96 (0.0854) | Disability: 0.0005 |
| 2 | Disability + Individualization | 0.0024 | 0.0019 | Disability: F,1 = 2.60 (0.1071)  Script: F,1 = 11.50 (0.0007) | Disability: 0.0004  Script: 0.0019 |
| 3 | Disability + Individualization + Engagement | 0.1018 | 0.0994 | Disability: F,1 = 17.55 (<0.0001)  Script: F,1 = 5.33 (0.0209)  Engagement: F,2 = 329.98 (<0.0001) | Disability: 0.0029  Script: 0.0009  Engagement: 0.0995 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1018 | 0.0000 | Disability: F,1 = 17.61 (<0.0001)  Script: F,1 = 5.39 (0.0203)  Engagement: F,2 = 330.00 (<0.0001)  Material: F,1 = 0.18 (0.6732) | Disability: 0.0029  Script: 0.0009  Engagement: 0.0995  Material: 0.0000 |

### Appendix 9.D: Model Analysis Summaries for Intellectual Disabilities Versus Learning Disabilities

Table 9.D.1 Model Summary—Grade Five, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.1058 | N/A | Disability: F,1 = 318.52 (<0.0001) | Disability: 0.1058 |
| 2 | Disability + Individualization | 0.1059 | 0.0001 | Disability: F,1 =318.62 (<0.0001)  Script: F,1 = 0.27 (0.6043) | Disability: 0.1059  Script: 0.0001 |
| 3 | Disability + Individualization + Engagement | 0.2247 | 0.1188 | Disability: F,1 = 263.81 (<0.0001)  Script: F,1 = 2.73 (0.0984)  Engagement: F,2 = 194.66 (<0.0001) | Disability: 0.0901  Script: 0.0010  Engagement: 0.1275 |
| 4 | Disability + Individualization + Engagement + Materials | 0.2247 | 0.0000 | Disability: F,1 = 263.69 (<0.0001)  Script: F,1 = 2.72 (0.0.0995)  Engagement: F,2 = 194.30 (<0.0001)  Material: F,1 = 0.00 (0.9714) | Disability: 0.0901  Script: 0.0010  Engagement: 0.1274  Material: 0.0000 |

Table 9.D.2 Model Summary—Grade Eight, Physical Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0768 | N/A | Disability: F,1 = 241.37 (<0.0001) | Disability: 0.0768 |
| 2 | Disability + Individualization | 0.0770 | 0.0002 | Disability: F,1 = 240.09 (<0.0001)  Script: F,1 = 0.51 (0.4757) | Disability: 0.0765  Script: 0.0002 |
| 3 | Disability + Individualization + Engagement | 0.2033 | 0.1263 | Disability: F,1 = 174.98 (<0.0001)  Script: F,1 =0.01 (0.9253)  Engagement: F,2 = 225.39 (<0.0001) | Disability: 0.0575  Script: 0.0000  Engagement: 0.1358 |
| 4 | Disability + Individualization + Engagement + Materials | 0.2033 | 0.0000 | Disability: F,1 = 174.76 (<0.0001)  Script: F,1 =0.01 (0.9275)  Engagement: F,2 = 224.46 (<0.0001)  Material: F,1 = 0.01 (0.9079) | Disability: 0.0574  Script: 0.0000  Engagement: 0.1353  Material: 0.0000 |

Table 9.D.3 Model Summary—Grade Eight, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0784 | N/A | Disability: F,1 = 246.57 (<0.0001) | Disability: 0.0784 |
| 2 | Disability + Individualization | 0.0788 | 0.0004 | Disability: F,1 = 244.67 (<0.0001)  Script: F,1 = 1.43 (0.2326) | Disability: 0788  Script: 0.0005 |
| 3 | Disability + Individualization + Engagement | 0.1604 | 0.0816 | Disability: F,1 = 184.54 (<0.0001)  Script: F,1 = 0.43 (0.5128)  Engagement: F,2 = 138.33 (<0.0001) | Disability: 0.0604  Script: 0.0001  Engagement: 0.0880 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1604 | 0.0000 | Disability: F,1 =184.20 (<0.0001)  Script: F,1 = 0.44 (0.5072)  Engagement: F,2 = 138.07 (<0.0001)  Material: F,1 = 0.04 (0.8433) | Disability: 0.0604  Script: 0.0002  Engagement: 0.0878  Material: 0.0000 |

Table 9.D.4 Model Summary—High School, Life Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0496 | N/A | Disability: F,1 = 248.00 (<0.0001) | Disability: 0.0496 |
| 2 | Disability + Individualization | 0.0496 | 0 | Disability: F,1 = 247.82 (<0.0001)  Script: F,1 = 0.43 (0.5110) | Disability: 0.0496  Script: 0.0001 |
| 3 | Disability + Individualization + Engagement | 0.1242 | 0.0746 | Disability: F,1 = 213.56 (<0.0001)  Script: F,1 = 0.34 (0.5586)  Engagement: F,2 = 200.13 (<0.0001) | Disability: 0.0436  Script: 0.0001  Engagement: 0.0786 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1243 | 0.0001 | Disability: F,1 = 214.04 (<0.0001)  Script: F,1 = 0.27 (0.6040)  Engagement: F,2 = 200.41 (<0.0001)  Material: F,1 = 0.61 (0.4349) | Disability: 0.0437  Script: 0.0001  Engagement: 0.0788  Material: 0.0001 |

Table 9.D.5 Model Summary—High School, Physical Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0382 | N/A | Disability: F,1 = 188.74 (<0.0001) | Disability: 0.0382 |
| 2 | Disability + Individualization | 0.0392 | 0.0010 | Disability: F,1 =187.82 (<0.0001)  Script: F,1 = 4.92 (0.0267) | Disability: 0.0380  Script: 0.0010 |
| 3 | Disability + Individualization + Engagement | 0.1668 | 0.1276 | Disability: F,1 = 159.72 (<0.0001)  Script: F,1 = 1.06 (0.3023)  Engagement: F,2 = 359.23 (<0.0001) | Disability: 0.0329  Script: 0.0002  Engagement: 0.1328 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1671 | 0.0003 | Disability: F,1 = 160.43 (<0.0001)  Script: F,1 = 1.16 (0.2816)  Engagement: F,2 = 360.10 (<0.0001)  Material: F,1 = 1.65 (0.1987) | Disability: 0.0331  Script: 0.0002  Engagement: 0.1331  Material: 0.0004 |

Table 9.D.6 Model Summary—High School, Physical Sciences, Activity 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0414 | N/A | Disability: F,1 = 205.25 (<0.0001) | Disability: 0.0414 |
| 2 | Disability + Individualization | 0.0425 | 0.0011 | Disability: F,1 = 203.65 (<0.0001)  Script: F,1 = 5.25 (0.0219) | Disability: 0.0411  Script: 0.0011 |
| 3 | Disability + Individualization + Engagement | 0.1900 | 0.1475 | Disability: F,1 = 173.39 (<0.0001)  Script: F,1 = 0.58 (0.4455)  Engagement: F,2 = 428.42 (<0.0001) | Disability: 0.0357  Script: 0.0001  Engagement: 0.1545 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1900 | 0.0000 | Disability: F,1 = 173.59 (<0.0001)  Script: F,1 = 0.59 (0.4443)  Engagement: F,2 = 426.83 (<0.0001)  Material: F,1 = 0.30 (0.5852) | Disability: 0.0357  Script: 0.0001  Engagement: 0.1540  Material: 0.0001 |

Table 9.D.7 Model Summary—High School, Earth and Space Sciences, Activity 1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| No. | Model | R2 | Difference in R2 | Significance Tests | Partial η2 |
| 1 | Disability | 0.0364 | N/A | Disability: F,1 = 179.56 (<0.0001) | Disability: 0.0364 |
| 2 | Disability + Individualization | 0.0392 | 0.0028 | Disability: F,1 = 179.63 (<0.0001)  Script: F,1 = 13.89 (0.0002) | Disability: 0.0364  Script: 0.0029 |
| 3 | Disability + Individualization + Engagement | 0.1297 | 0.0905 | Disability: F,1 = 146.11 (<0.0001)  Script: F,1 = 8.61 (0.0034)  Engagement: F,2 = 242.94 (<0.0001) | Disability: 0.0302  Script: 0.0018  Engagement: 0.0937 |
| 4 | Disability + Individualization + Engagement + Materials | 0.1300 | 0.0003 | Disability: F,1 =146.85 (<0.0001)  Script: F,1 = 8.82 (0.0030)  Engagement: F,2 = 243.40 (<0.0001)  Material: F,1 = 1.43 (0.2320) | Disability: 0.0303  Script: 0.0019  Engagement: 0.0939  Material: 0.0003 |

## Quality Control Procedures

The California Department of Education (CDE) and Educational Testing Service (ETS) implemented rigorous quality control procedures throughout the test development, administration, scoring, analyses, and reporting processes. As part of this effort, ETS California Assessment of Student Performance and Progress (CAASPP) program staff worked with the ETS 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 Embedded PT Development

ETS’ goal is to provide the best standards-based embedded performance tasks (PTs) for the California Alternate Assessment (CAA) for Science. Embedded PTs developed for the CAA for Science underwent an extensive embedded PT review process. The item writers hired to develop CAA items were trained in CAASPP and ETS policies on quality control of item content, sensitivity, and bias guidelines, as well as on guidelines for accessibility, to ensure that the items allow the widest possible range of students to demonstrate their content knowledge.

Once a written item was accepted for authoring—that is, once it was entered into ETS’ item bank and formatted for use in an assessment—ETS employed a series of internal and external reviews. These reviews used established criteria and specifications to judge the quality of items and to ensure that each item measured what it was intended to measure. These reviews also examined the overall quality of the test items before they were presented to the CDE and item reviewers. Finally, a group of California educators reviewed the items for accessibility, bias and sensitivity, and content prior to their administration to students. The details on quality control of item development are described in subsection [*3.2 Embedded PT and Item Review Process*](#_Embedded_PT_and_1).

### Quality Control of Test Assembly and Delivery

The assembly of all test forms must conform to the mutually agreed-upon test design that represents a set of constraints and specifications. These constraints are critical to the formation of valid assessments. The blueprint for the CAA for Science was approved in January 2018 and test forms were assembled following the approved blueprint (CDE, 2018).

#### Quality Control of Test Form Development

ETS conducted multiple levels of quality assurance checks on each constructed test form to ensure it met defined statistical criteria. For the 2018–2019 CAA for Science, both ETS Assessment & Learning Technology Development (ALTD) and psychometric staff reviewed and signed off on the accuracy of forms before the test forms were put into production for the field test administration. Detailed information related to test assembly can be found in subsection [*3.2.1 Selection of Embedded PTs and Items*](#_Selection_of_Embedded).

In particular, the assembly of all test forms went through a certification process that included various checks to verify that

* all correct answers were correct;
* items were scored correctly in the item bank;
* all embedded PTs aligned with the standard;
* all content in the embedded PT was correct;
* 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 art was correct;
* there were no errors in spelling or grammar; and
* embedded PTs adhered to the approved style guide.

Reviews were also conducted for functionality and sequencing during the user acceptance testing (UAT) process to ensure all items functioned as expected.

#### Quality Control of Test Assignment

Test assignment for the CAASPP assessments, including the CAA for Science, is controlled by the Test Operations Management System (TOMS) using student demographic information received from the California Longitudinal Pupil Achievement Data System (CALPADS) (CDE, 2019). The two systems are kept in sync during the testing window.

Students at eligible grade levels were assigned to the Smarter Balanced assessments (in grades three through eight and grade eleven) and the California Science Test (CAST) (grades five and eight and high school) by default. For students eligible for the CAA for Science—that is, grades five, eight, and high school (grade ten, eleven, or twelve)—local educational agencies (LEAs) logged on to TOMS and assigned students to take the alternate assessment, which automatically unassigned those students from taking the CAST.

The quality of test assignment for the CAA for Science was monitored and controlled through several strategies. TOMS enforced preconditions for eligibility for the CAAs by permitting assignment only for students with an Individuals with Disabilities Education Act[[8]](#footnote-9) indicator of “Yes” in TOMS. This indicator is set to “Yes” when the CALPADS *Education Program* field (field 3.13) is equal to 144 (Special Education) and the primary disability code (field 3.21) is not set to blank.

Additionally, TOMS prevented the prohibited “mixing and matching” of assessments. For example, a student assigned to take an alternate assessment for any content area will automatically be prevented from assignment to a general assessment for another content area.

#### Quality Control of Test Administration

The quality of test administration is managed through comprehensive rules and guidelines for maintaining the security and standardization of CAASPP assessments, including the CAA for Science. LEAs received training on these topics and were provided with tools to report security incidents and resolve testing discrepancies for specific testing sessions.

Several strategies are utilized to monitor and control the quality of test administration for the CAA for Science as well as all assessments administered as part of the CAASPP System. A fully staffed support center, the California Technical Assistance Center (CalTAC), supports all LEAs in the administration of all CAASPP assessments. CalTAC is guided by a core group of LEA outreach and advocacy staff who manage communications to LEAs, regional and web-based trainings, and the CAASPP website, which houses a full range of manuals, videos, and other instructional and support materials. In addition to providing guidance and answering questions, CalTAC regularly conducts outreach campaigns on particular administration topics to ensure all LEAs understand correct test administration procedures.

The ETS Office of Testing Integrity (OTI) reinforces the quality control procedures for test administration, providing quality assurance services for all testing programs managed by ETS. The OTI’s detailed quality control procedures are described in subsection [*5.7.1. ETS’ Office of Testing Integrity*](#_ETS’_Office_of).

#### Quality Control of Machine-Scoring Procedures

To ensure valid item-level scoring for the CAA for Science, quality control procedures were employed by American Institutes for Research (AIR), the CAASPP subcontractor responsible for providing the CAASPP test delivery system (TDS) and scoring machine-scorable items. A final comparison of the test map to each online 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 quality assurance (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 had 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 for reporting was pulled and transmitted to ETS in a predetermined results format.

### Quality Control of Test Materials

The steps taken to develop and ensure the quality of the online assessments are described in [*Chapter 3: Embedded PT Development and Review*](#_Embedded_PT_Development).

#### Test Administration Manuals

ETS staff consulted with internal subject matter experts and conducted validation checks to verify that test directions and administration manuals accurately matched the test materials and testing processes. Copy editors and content editors reviewed each document for spelling, grammar, accuracy, and adherence to CDE style. Each document was required to be approved by the CDE before it could be published to the CAASPP website. Only nonsecure documents were posted to this website. Secure materials, such as the *CAA for Science Embedded Performance Task Directions for Administration*, were made available to designated LEA staff through TOMS, which required a secure logon.

The manuals used in the administration of the CAA are listed in subsection [*5.4.4 Instructions for Test Examiners and Staff Involved in CAA for Science Administration*](#_Instructions_for_Test).

#### Processing Test Materials

The following information was entered into the TDS by test examiners and transmitted from AIR to ETS each day:

* Student’s first name
* Statewide Student Identifier
* Results of the Student Response Check
* Any individualized scripts and materials used
* Responses for each item
* Results of the student survey
* Results of the student engagement survey
* Final score calculated by the test examiner

The AIR and ETS systems checked for the completeness of the student record and stopped records identified as having an error.

### Quality Control of Psychometric Processes

#### Development of Scoring Specifications

ETS scoring specifications for the CAA for Science 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 calculations for computing summary data.

#### Development of Scoring Procedures

ETS’ Enterprise Score Key Management System (eSKM) utilized scoring procedures specified by psychometricians and provides scoring services. ETS psychometricians carried out a series of quality control checks after scoring to ensure the accuracy of each score.

##### Enterprise Score Key Management System (eSKM) Processing

Prior to the test administration, ETS ALTD staff reviewed and verified the keys for all items. Then, these keys were provided to AIR for its machine-scoring implementation. After AIR finished machine-scoring, those scores and responses were delivered to ETS. AIR quality control of the machine-scoring procedure is described in subsection [*10.2.3 Quality Control of Test Administration*](#_Quality_Control_of_1).

ETS’ Centralized Repository Distribution System and Enterprise Service Bus departments collected and parsed .xml files that contained student response data from AIR. ETS’ eSKM system collected and calculated individual students’ overall scores (total raw scores) and generated student scores in the approved statistical extract format. These data extracts were sent to ETS’ Data Quality Services for data validation. Following successful validation, the student response statistical extracts were made available to the psychometric team.

ETS developed two parallel scoring systems to produce and verify student scores: the eSKM scoring system received an individual student’s item scores and item responses from AIR and calculated individual student scores for ETS’ reporting systems. The Psychometric Analysis & Research team also computed individual student scores based on item scores delivered by AIR. The scores from the two sources were then compared for internal quality control. Any differences in the scores were discussed and resolved. All scores complied with the ETS scoring specifications and passed the parallel scoring process to ensure the quality and accuracy of scoring and to support the transfer of scores into TOMS, the database of the student records scoring system.

##### Psychometric Processing

The psychometric analyses conducted at ETS underwent comprehensive quality checks by a team of psychometricians and data analysts. The ETS psychometric team reviewed the data files before conducting the statistical analyses to ensure the quality of the data. The team developed detailed checklists for each of the statistical procedures performed on each CAA for Science grade-level assessment. The classical item analyses and differential item functioning analyses were run by one data analyst and checked by a second data analyst. Results were then reviewed by the psychometricians to compile a list of flagged items for ETS ALTD staff for review. ALTD comments were reviewed by the psychometricians before items were approved for inclusion in additional analyses and before the data review meetings with the CDE.

### Quality Control of Reporting

An aggregate report summarizing the results of the 2018–2019 field test administration for the CAA for Science was provided to the CDE. The aggregate file contained the CAA for Science scores for a given grade that were aggregated at the school, LEA or direct funded charter school, county, and state levels. To ensure the quality of the aggregate file, two members of the ETS psychometrics team individually reviewed the files and worked with ETS Information Technology to resolve any discrepancies before the files were posted to the CDE.

### References

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacramento, CA: California Department of Education.

California Department of Education. (2019). *TOMS pre-administration guide for CAASPP testing*. Sacramento, CA: California Department of Education.

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

## Continuous and Systematic Improvements

### Improvements from the Second-Year Pilot

There were several changes made to the California Alternate Assessment (CAA) for Science for the field test administration, on the basis of the lessons learned from the second-year pilot administered in 2017–2018 and on California’s adoption of a test blueprint in January 2018 that guides ongoing development and administration of the CAA for Science.

#### Changes to Test Administration

The following is a summary of improvements that Educational Testing Service, in coordination with the California Department of Education, made to the CAA for Science from the 2017–2018 second-year pilot to the 2018–2019 field test administration:

* Summaries of the content to be assessed were separated from each embedded performance task (PT) and centrally housed in grade-specific *Administration Planning Guides* that test examiners could access prior to testing in order to better plan their testing schedule.
* The test was administered online, which meant the Answer Recording Document was eliminated and test examiners no longer had to transcribe student responses into the Data Entry Interface. This was made possible by the conversion of test content from paper-delivered using electronic PDF files to online-delivered using items administered through the California Assessment of Student Performance and Progress test delivery system. A separate *Directions for Administration (DFA)* document provided guidelines and scripts for the test examiner to use.
* A new embedded PT structure was established that assesses two Science Connectors in each embedded PT, with a standardized number of items by complexity level and score points.
* The naming convention “orienting activity” was adopted and the presentation of the orienting activity was reorganized for each Science Connector set, with refined guidelines for individualization options on the orienting activities and some items.
* The general student survey—two survey questions that were asked of the student following each embedded PT—was eliminated, although the student engagement survey remained.

#### Changes to Content

There have been two pilots of the assessment and both were invaluable in gathering information to inform the structure of the field test and subsequent operational test.

##### Online Test Delivery

To support gathering information about how best to incorporate simple hands-on experiences as part of the assessment, the pilots were paper based, with the goal of transitioning to online for the field test. Thus, Pilot 2 in 2017–2018 was paper based, with teachers entering student responses into the online scoring system, while the 2018–2019 field test was administered completely online, as were the alternate assessments for English language arts/literacy (ELA) and mathematics.

In addition, with the launch of the online field test, the *DFA* that supports the administration of the assessment was then revised to have a similar structure and terminology as that of the *DFAs* for the ELA and mathematics assessments. This improves the experience of both the test examiners and the students to have that consistency across the three assessments.

The move to an online administration also made the use of color graphics possible. When the assessment was paper based, the variability in printer quality made it necessary to use black and white to ensure sufficient quality and clarity of the graphics.

##### Use of the Approved Blueprint in Test Design

The field test year was the first year to implement the test blueprint (CDE, 2018). Previous pilot administrations had varying numbers of items, with the goal of gathering information about the outcomes of different approaches.

##### Development of the *Administration Planning Guides*

Another improvement was the development of the *Administration Planning Guides* (CDE, 2019). These guides provided information that helped teachers plan for the next year’s assessment. The guides also provided information about the assessment and links to resources.

Updated guides will be published in May and provide information about the Science Connectors that will be targeted for assessment in the upcoming school year.

##### Simplification of Activities

Feedback from the field during the two pilots provided guidance in simplifying the activities that are part of the assessment. This includes both the orienting activities and the activities that may be embedded in the assessment as part of the items.

One natural outcome of this simplification was a reduction in the amount and variety of the materials needed, which then led to most teachers using the materials named in the activity directions instead of substituting alternate materials based on student needs.

### References

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacrament, CA: California Department of Education.

California Department of Education. (2019). *2018–19 CAA for Science administration planning guide: Grade eight.* Sacramento, CA: California Department of Education.

1. The total population of students with the most significant cognitive disabilities in the California kindergarten through grade twelve (K–12) public school system is approximately 38,000 (1 percent of the total student enrollment, which is provided in the CDE’s DataQuest website, for the 2015–2016 school year). [↑](#footnote-ref-2)
2. Retrieved from the CDE Fingertip Facts on Education in California – *CalEdFacts* web page [↑](#footnote-ref-3)
3. From the CDE California Longitudinal Pupil Achievement Data System (CALPADS) web page [↑](#footnote-ref-4)
4. This technical report is based on the version of Matrix One that was available during the 2018–2019 CAASPP administration. [↑](#footnote-ref-5)
5. Disability information was changed or removed after student testing. [↑](#footnote-ref-6)
6. S. 1177—114th Congress: Every Student Succeeds Act. 2015. Title 1, Part A, Subpart 1, Section 1111(b)(2)(D)(ii)(I) [↑](#footnote-ref-7)
7. The SAS PROC GLM CLASS statement will convert categorial independent variables into binary dummy variables (SAS Institute Inc, 2015). Users can use the CLASS statement to indicate which independent variables in the model are categorical and set a reference level for each categorical variable. [↑](#footnote-ref-8)
8. The Individuals with Disabilities Education Act is the primary federal program that authorizes state and local aid for special education and related services for children with disabilities. [↑](#footnote-ref-9)