

California Department of Education Assessment Development & Administration Division



California Assessment of Student Performance and Progress

# California Assessment of Student Performance and Progress

# California Alternate Assessment for Science Second-Year Pilot Technical Report

# 2017–18 Administration

Final Submitted June 20, 2020 Educational Testing Service



Contract No. CN150012

# **Table of Contents**

Chapter 1: Introduction	1
1.1. Background	1
1.2. Purpose of the Second-Year Pilot	2
1.3. Second-Year Pilot Content	2
1.3.1. Assessment Model	2
1.3.2. California Next Generation Science Standards Core Content Connectors (Science Standards Core Science Standards Core Content Connectors (Science Standards Core Content Connectors (Science Standards Core Science Science Standards Core Science Scien	<u>sience</u>
Connectors)	<u>3</u> 3
1.3.3. Test Components for the Second-Year Pilot	
1.4. Intended Population	4
1.5. Testing Window and Times	5
1.6. Limitations of the Assessment	5
1.7. Groups and Organizations Involved with the Assessment	5
1.7.1. State Board of Education (SBE)	5
1.7.2. California Department of Education (CDE)	6
1.7.3. California Educators	6
1.7.4. Contractors	6
1.8. Systems Overview and Functionality	7
1.8.1. Test Operations Management System (TOMS)	7
1.8.2. Online Reporting System (ORS)	7
1.9. Overview of the Technical Report	7
References	9
Chapter 2: Overview of the CAA for Science Second-Year Pilot Processes	<u> </u>
2.1. Embedded Performance Task (PT) Development and Review	10
2.1.1. Selection of Science Connectors for Embedded PT Development	10
2.1.2. Universal Design Principles	10
2.1.3. Embedded PTs Developed for Grades Five and Eight and High School	11
2.2. Test Administration	11
2.2.1. Accessibility Features	11
2.2.2. Individualizations	11
2.3. Participation	11
2.4. Scores	12
2.4.1. Score Reporting	12
2.4.2. Aggregation Procedures	12
2.5. Overview of Psychometric Analyses	13
2.5.1. Analysis Types	13
2.5.2. Description of DIF Analyses	13
References	14
Appendix 2.A Participation Rates	15
Chapter 3: Embedded Performance Task (PT) Development and Review	19
3.1. Embedded PT Development	19
3.1.1. Specifications for the Embedded PTs	19
3.1.2. Embedded PT Format	20
3.1.3. Recruitment and Selection of Embedded PT Item Writers	21
3.1.4. Embedded PT Item Writer Training	21
3.2. Embedded PT Review Process	22

3.2.1. Selection of Embedded PTs	22
3.2.2. ETS Content Review	22
3.2.3. ETS Editorial Review	23
3.2.4. ETS Sensitivity and Fairness Review	23 23
3.3. California Educator's Review	23
3.3.1. California Educators as Content Experts	23
3.3.2. Composition of Item Review Panels	23
3.3.3. Meetings for Review of CAA for Science Items	25
3.4. Data Review	26
Chapter 4: Test Administration	27
4.1. Grade Assignment for High Schools	27
4.2. Administration Preparations	27
4.2.1. Guides and Videos	27
4.2.2. Practice and Training Tests	28
4.3. Test Administration	28
4.3.1. Administration of the Student Response Check (SRC)	28
4.3.2. Administration of the Embedded PTs	28
4.3.3. Administration of the Survey	29
4.4. Procedures to Maintain Standardization	29
4.4.1. LEA CAASPP Coordinator	29
4.4.2. CAASPP Test Site Coordinator	<u>30</u>
4.4.3. Test Examiners	31
4.4.4. Instructions for Test Examiners and Staff Involved in CAA Administration	<u>31</u>
4.5. Accessibility Features for the Second-Year Pilot	32
4.5.1. Individualizations	32
4.5.2. Choice of Administration Scripts	36
4.5.3. Choice of Materials	36
4.5.4. Type and Level of Accommodations	36
4.6. Test Security and Confidentiality	36
4.6.1. ETS' Office of Testing Integrity (OTI)	36
4.6.2. Procedures to Maintain Standardization of Test Security	37
4.6.3. Security of Electronic Files Using a Firewall	37
4.6.4. Transfer of Scores via Secure Data Exchange	37
4.6.5. Data Management in the Secure Database 4.6.6. Statistical Analysis on Secure Servers	<u>38</u> 38
4.6.7. Student Confidentiality	38
4.6.8. Student Test Results	38
4.6.9. Security and Test Administration Incident Reporting System (STAIRS) Process	40
References	42
Chapter 5: Scoring and Reporting	43
5.1. CAA for Science Scoring Process	<u>43</u>
5.2. Types of Scores	43
5.2.1. Percent Correct	43
5.2.2. Preliminary Indicator Categories	43
5.2.3. Aggregate Score Reporting	44
5.3. Survey Questions Regarding Test Administration	48
5.3.1. Student Engagement	48
	-+0

5.3.2. Individualization of the Test	49
References	51
Appendix 5.A: Demographic Summaries	52
Chapter 6: Analyses	<u>62</u>
6.1. Sample Used for the Analyses	62
6.2. Classical Item Analyses	62
6.2.1. Classical Item Difficulty Indices (p-value and Average Item Score)	62
6.2.2. Item Discrimination (Item-Total Correlation)	63
6.2.3. Distribution of Item Scores	64
6.2.4. Summary of Classical Item Analysis Flagging Criteria	<u>64</u>
6.2.5. Classical Item Analysis Results Summary	<u>64</u>
6.3. Test Completion	<u>65</u>
6.4. Task Difficulty (Overall and by Embedded PT)	<u>65</u>
6.5. Differential Item Functioning (DIF)	<u>66</u>
6.5.1. Dichotomous Items	<u>66</u>
6.5.2. Polytomous Items	67
6.5.3. DIF Categories and Definitions	<u>68</u>
6.5.4. Items Exhibiting Significant DIF	<u>69</u>
References	70
Appendix 6.A: Classical Item Analyses	71
Appendix 6.B: Completion Rates	75
Appendix 6.C: Distribution of Raw Scores Overall and by Embedded Performance Task	
Annendix 6 Dr Differential Item Eurotianing (DIE) Analysis: Number and Persentage of	<u>76</u>
Appendix 6.D: Differential Item Functioning (DIF) Analysis: Number and Percentage of Items in Each DIF Category	79
Chapter 7: Surveys	<u>73</u> 84
7.1. Survey Design and Questionnaire Development	84
7.1.1. Student Survey and Student Engagement Survey	84
7.1.2. Test Examiner Survey	84
7.2. Student Survey Administration	85
7.3. Student Survey Results	86
7.3.1. Summary of Survey Results Regarding Student Experiences	86
7.4. Test Examiner Survey Results	91
7.4.1. Responses to Background Questions	91
7.4.2. Responses Regarding Testing Time	93
7.4.3. Responses Regarding Individualization	94
7.4.4. Responses Regarding Clarity of Instructions Provided	95
7.4.5. Open-ended Responses	95
Chapter 8: Embedded Performance Task and Test Comparability Considerations	97
8.1. Considerations	97
8.2. Summary of Findings from the Choice of Materials and Individualization Analysis	99
8.2.1. Individualization Analysis	99
8.2.2. Model Analysis	99
8.2.3. Results of the Individualization and Model Analyses	100
8.3. Implications for the Field Test	
	<u>101</u>
8.3.1. Key Findings	<u>101</u> 101

References	102
Appendix 8.A: Choice of Materials and Individualization Analysis Data	103
Appendix 8.B: Model Analysis Summaries	106
Chapter 9: Quality Control Procedures	112
9.1. Quality Control of Embedded PT Development	112
9.2. Quality Control of Test Assembly and Delivery	112
9.2.1. Quality Control of Test Form Development	112
9.2.2. Quality Control of Test Assignment	113
9.2.3. Quality Control of Test Administration	114
9.3. Quality Control of Test Materials	114
9.3.1. Test Administration Manuals	114
9.3.2. Processing Test Materials	114
9.4. Quality Control of Psychometric Processes	115
9.4.1. Development of Scoring Specifications	115
9.4.2. Development of Scoring Procedures	115
9.4.3. Psychometric Processing	115
9.5. Quality Control of Reporting	115
References	117
Chapter 10: Continuous and Systematic Improvements	118
10.1. Improvements from the First-Year Pilot	<u> </u>
10.1.1. Changes to Test Administration	<u> </u>
10.1.2. Answer Recording Document	<u> </u>
10.1.3. Student Independence Evaluations	119
10.1.4. Changes to Content	119
10.2. In-Person Observations	119
10.2.1. Observation Protocols	119
10.2.2. Summary of the In-Person Observations	120
10.2.3. Lessons Learned from the In-Person Observations	120
10.2.4. Postobservation Interviews	121
10.3. Implications Based on the Survey Results	122
10.3.1. Test Delivery	123
10.3.2. Use of Graphics	123
10.3.3. Disseminating Information About the CAA for Science	123

# List of Tables

Acronyms and Initialisms Used in the California Alternate Assessment for Science	
Technical Report	Viii
Table 1.1 Organization of the Science Connectors	3
Table 2.A.1 CAA for Science Second-Year Pilot Participation Rates—Enrolled Students	<u>15</u>
Table 2.A.2 CAA for Science Second-Year Pilot Participation Rates for Grade Five by	
Student Group	<u>16</u>
Table 2.A.3 CAA for Science Second-Year Pilot Participation Rates for Grade Eight by	
Student Group	17
Table 2.A.4 CAA for Science Second-Year Pilot Participation Rates for High School by	
Student Group	<u>18</u>

Table 3.1 Number of Items and Points for Each PT	21
Table 3.2 Number of Item Reviewers with Each Qualification	24
Table 4.1 Individualizations—Grade Five	33
Table 4.2 Individualizations—Grade Eight	<u>34</u>
Table 4.3 Individualizations—High School	<u>35</u>
Table 5.1 Indicator Categories	44
Table 5.2 Cut Scores for Preliminary Categories	44
Table 5.3 Grade Five Preliminary Indicator Conversion Table	<u>45</u>
Table 5.4 Grade Eight Preliminary Indicator Conversion Table	<u>45</u>
Table 5.5 High School Preliminary Indicator Conversion Table	<u>46</u>
Table 5.6 Demographic Student Groups to Be Reported	47
Table 5.A.1 Demographic Summary for Grade Five	<u>52</u>
Table 5.A.2 Demographic Summary for Grade Eight	<u>56</u>
Table 5.A.3 Demographic Summary for High School	<u>59</u>
Table 6.1 Classical Item Statistics	<u>65</u>
Table 6.2 Raw Score Summary for Each Embedded PT	<u>65</u>
Table 6.3 DIF Categories for Dichotomous Items	<u>68</u>
Table 6.4 DIF Categories for Polytomous Items	<u>68</u>
Table 6.5 Student Groups for DIF Comparison	<u>69</u>
Table 6.A.1 Classical Item Statistics for Each Embedded PT	71
Table 6.A.2 Average Item Score and Polyserial Correlation for Each Item: Grade Five	72
Table 6.A.3 Average Item Score and Polyserial Correlation for Each Item: Grade Eight	72
Table 6.A.4 Average Item Score and Polyserial Correlation for Each Item: High School	73
Table 6.A.5 Distribution of Item Scores for the Polytomous Item: Grade Five	73
Table 6.A.6 Distribution of Item Scores for the Polytomous Items: Grade Eight	74
Table 6.A.7 Distribution of Item Scores for the Polytomous Item: Grade High School	74
Table 6.B.1 Completion Rate	75
Table 6.B.2 Percentage of Students Completing the PTs	75
Table 6.C.1 Distribution of Students for Total Score and PT Scores for Grade Five	<u>76</u>
Table 6.C.2 Distribution of Students for Total Score and PT Scores for Grade Eight	77
Table 6.C.3 Distribution of Students for Total Score and PT Scores for High School	<u>78</u>
Table 6.D.1 DIF for Grade Five	79
Table 6.D.2 DIF for Grade Five (Continued)	<u>80</u>
Table 6.D.3 DIF for Grade Eight	<u>80</u>
Table 6.D.4 DIF for Grade Eight (Continued)	<u>81</u>
Table 6.D.5 DIF for High School	<u>81</u>
Table 6.D.6 DIF for High School (Continued)	<u>82</u>
Table 6.D.7 DIF for High School (Continued)	<u>82</u>
Table 6.D.8 Item Exhibiting Significant DIF—Grade Five	<u>83</u>
Table 7.1 Student Survey Responses—How did you feel about taking this test?	<u>86</u>
Table 7.2 Student Survey Responses—Did you have enough time to complete the test?	<u>86</u>
Table 7.3 Student Engagement Survey Responses—Select the mode(s) of	
communication used by the student on this performance task. (Select all that apply)	87
Table 7.4 Student Engagement Survey Responses—"Other" Mode(s) of Communication:	
Most Common Responses for Grade Five	<u>89</u>
Table 7.5 Student Engagement Survey Responses—"Other" Mode(s) of Communication:	
Most Common Responses for Grade Eight	<u>90</u>

Table 7.6 Student Engagement Survey Responses—"Other" Mode(s) of Communication:	
Common Responses for High School	90
Table 7.7 Student Engagement Survey Responses—How engaged was the student with	
this performance task you just administered?	91
Table 7.8 Teacher for Student Tested	92
Table 7.9 How Many Students Tested Per Test Examiner	92
Table 7.10 Grades Administered	92
Table 7.11 Testing Weeks Approach	93
Table 7.12 Cumulative Testing Time for All Three PTs Slider Question	<u>93</u>
Table 7.13 Frequency of Individualization	94
Table 7.14 Importance of Individualization to Student Performance	94
Table 7.15 Elements Individualized	<u>95</u>
Table 7.16 Clarity of DFA	<u>95</u>
Table 8.1 Linear Models Estimated in Material Choices Analysis	<u>98</u>
Table 8.A.1 Individualizations—Grade Five	103
Table 8.A.2 Individualizations—Grade Eight	104
Table 8.A.3 Individualizations—High School	<u>105</u>
Table 8.B.1 Model Summary—Grade Five, PT 1, Activity 1	<u>106</u>
Table 8.B.2 Model Summary—Grade Five, PT 2, Activity 1	<u>106</u>
Table 8.B.3 Model Summary—Grade Five, PT 3, Activity 2	107
Table 8.B.4 Model Summary—Grade Eight, PT 1, Activity 1	107
Table 8.B.5 Model Summary—Grade Eight, PT 1, Activity 2	<u>108</u>
Table 8.B.6 Model Summary—Grade Eight, PT 2, Activity 1	<u>108</u>
Table 8.B.7 Model Summary—Grade Eight, PT 2, Activity 2	<u>108</u>
Table 8.B.8 Model Summary—Grade Eight, PT 3, Activity 1	109
Table 8.B.9 Model Summary—Grade Eight, PT 3, Activity 2	109
Table 8.B.10 Model Summary—High School, PT 1, Activity 1	109
Table 8.B.11 Model Summary—High School, PT 1, Activity 2	110
Table 8.B.12 Model Summary—High School, PT 2, Activity 1	110
Table 8.B.13 Model Summary—High School, PT 2, Activity 2	111
Table 8.B.14 Model Summary—High School, PT 3, Activity 1	111
Table 10.1 Number of Student Participants and Test Examiners by Grade	120
Table 10.2 Number of Observations by Grade and Embedded PT	120

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

Term	Definition
AD	Assessment Development American Institutes for Research
AIR	
AIS	average item score
ASL	American Sign Language
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	<u> </u>
CCSS	Common Core State Standards
CDE	California Department of Education
DEI	Data Entry Interface
DFA	Directions for Administration
DIF	differential item functioning
EC	Education Code
ELA	English language arts/literacy
ETS	Educational Testing Service
EUs	essential understandings
FKSA	focal knowledge, skills, and abilities
IEP	individualized education program
LEA	local educational agency
MH DIF	Mantel-Haenszel differential item functioning
ORS	Online Reporting System
OTI	Office of Testing Integrity
PE	performance expectation
PT	performance task
SBE	State Board of Education
SD	standard deviation
Science Connectors	Science Core Content Connectors
SFTP	Secure File Transfer Protocol
SMD	standardized mean difference
SRC	Student Response Check
SS3	type III sum of squares
STAIRS	Security and Test Administration Incident Reporting System
TOMS	Test Operations Management System
USC	United States Code

# **Chapter 1: Introduction**

## 1.1. 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 Core Content Connectors (Science Connectors) derived from the CA NGSS. Its second pilot was administered during the 2017–18 CAASPP administration.

The CAA for Science is for students in grades five and eight and high school whose individualized education program (IEP) teams have determined that a student should take the CAA (California Department of Education [CDE], 2018a). 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.

During the 2017–18 administration, the overall CAASPP System had the following components:

- 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, which may be administered to students at any grade level
  - Digital Library—Tools, lesson plans, and practices 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, eight, and high school (grades ten, eleven, or twelve; these are the California Science Test [CAST] and the CAA for Science)
- A primary language assessment, the Standards-based Tests in Spanish for Reading/ Language Arts, in grades two through eleven (optional for eligible Spanish-speaking English learners)
- A new primary language assessment, the California Spanish Assessment, delivered in pilot form at selected local educational agencies (LEAs), to students in grades three through eight and high school who are Spanish-speaking English learners or students seeking a measure that recognizes their Spanish reading, writing, and listening skills

More background information about the CAASPP System can be found on the CAASPP Description – *CalEdFacts* web page at <u>http://www.cde.ca.gov/ta/tg/ai/cefcaaspp.asp</u>.

# **1.2. Purpose of the Second-Year Pilot**

The purpose of the second CAA for Science pilot was to permit a second round of information-gathering about the new 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 pilot built on the infrastructure developed for the first pilot by adding a systematic data collection component that allowed test examiners to enter results from administering the second-year pilot PTs to students into the online Data Entry Interface (DEI) so ETS could analyze the data. The overarching goal of the second-year pilot was to lay 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.

## **1.3. Second-Year Pilot Content**

### 1.3.1. Assessment Model

The California State Board of Education (SBE) approved the conceptual design for the CAA for Science in July 2016. This assessment used an embedded PT design, meaning that each PT should be administered shortly after classroom instructional activities. Test examiners administer a set of test questions measuring two Science Connectors from one of the three science domains (CDE, 2018b).

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<sup>1</sup> also makes the use of this assessment model reasonable.

<sup>&</sup>lt;sup>1</sup> 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

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

The assessment is to be 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. <u>Table 1.1</u> summarizes the structure and 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 regards 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			

Table 1.1	<b>Organization of the Science Connectors</b>
-----------	---

### 1.3.3. Test Components for the Second-Year Pilot

The 2017–18 second-year pilot of the CAA for Science involved four components:

- 1. three embedded PTs,
- 2. a brief student survey,
- 3. an optional test examiner survey, and
- 4. an optional training PT.

#### 1.3.3.1 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.

For the 2017–18 CAASPP administration, three PTs were developed for the CAA for Science: one each for grade five, grade eight, and high school (i.e., grade ten, eleven, or twelve). Each PT included information for the test examiner, describing the hands-on activity and how to administer the PT items. The PT item types included selected-response and identification items; these are described in subsection <u>3.1.2 Embedded PT Format</u>.

<sup>38,000 (</sup>one percent of the total student enrollment, which is provided in the CDE's DataQuest website, for the 2015–16 school year. Data was retrieved from <a href="http://dq.cde.ca.gov/dataquest/">http://dq.cde.ca.gov/dataquest/</a>.

The secure embedded PTs were delivered to LEAs as downloadable PDFs within the Test Operations Management System (TOMS). Test examiners were instructed to print these documents, record student responses on the printed documents, and store student responses in a secure location. LEA CAASPP coordinators were directed to store student results locally for one year.

#### 1.3.3.2 Survey for Students

During the 2017–18 administration year, students responded 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 purposes of the student survey were as follows:

- 1. Provide electronic documentation of student participation
- 2. Collect basic information about students' experiences with the assessment process

This survey was included in the last section of the downloadable embedded PT document. After marking responses to the student surveys, test examiners entered each student's survey responses into the DEI. See <u>chapter 7</u> for additional information about the student survey design and results.

#### 1.3.3.3 Student Engagement Survey

As the final step in administering an embedded PT, the test examiner answered two questions regarding the student's engagement with the administered PT. This student engagement survey differs from the survey for students in that it is taken by the test administrator, while the survey for students is taken by students.

The student engagement survey was included in the last section of the downloadable embedded PT document. The student engagement survey asked the test examiner about the mode of communication used by the student, as well as the level of engagement of the student with the PT. After marking responses to the student engagement surveys, test examiners entered the survey results into the DEI. See <u>chapter 7</u> for additional information about the student engagement survey design and results.

#### 1.3.3.4 Optional Test Examiner Survey

An optional survey was presented to test examiners to obtain teachers' feedback on the pilot administration and assessment processes in order to guide the implementation of each respective assessment. This survey was linked on the CAASPP Portal at <a href="http://www.caaspp.org/">http://www.caaspp.org/</a> and hosted on SurveyGizmo.com, a website with survey-creation and hosting services. See <a href="http://chapter-7">chapter 7</a> for additional information about the optional test examiner survey and results.

#### 1.3.3.5 Optional Sample Performance Tasks

Test examiners had an opportunity to gain familiarity with the new assessment and embedded PT format through access to sample PTs that were linked on caaspp.org as downloadable PDF files. There was one sample PT for grade five and two each for grade eight and high school.

Like the embedded PTs, the sample PT contained an answer key that could be used to allow teachers or parents/guardians to score student responses.

## **1.4. Intended Population**

All eligible students enrolled in grades five, eight, and twelve 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. Only students whose parents/guardians submit a written request may opt out of taking the tests (*Education Code [EC]* Section 60615).

## 1.5. Testing Window and Times

For the 2017–18 CAASPP administration, the CAA for Science second-year pilot embedded PTs were available for administration on or after November 1, 2017. Test examiners downloaded and administered the embedded PTs one on one with students and then entered the results, survey responses, and test individualizations into the DEI starting on January 9, 2018, through the last day of instruction at the LEA or July 16, 2018, whichever came first (5 *CCR*, Section 855[a][2]).

Similar to other CAASPP assessments, the CAA for Science second-year pilot PTs were untimed for students. 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 PTs occured over as many days as required to meet a student's needs.

## **1.6. Limitations of the Assessment**

The CAA for Science second-year pilot 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 pilot tests was to evaluate the items rather than students' knowledge, skills, and abilities, the pilot tests were not a full representation of the assessment model for the CA NGSS Science Connectors.

Results for the CAA for Science were reported using preliminary indicators, which are descriptive statements with corresponding threshold scores. Preliminary indicators are general, rather than precise, indications of student content knowledge. Their purpose is to help LEAs during the transition period. Therefore, caution should be used when interpreting the preliminary indicator results.

# **1.7. Groups and Organizations Involved with the Assessment**

## 1.7.1. 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 the 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, 2017).

## 1.7.2. 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 10,450<sup>2</sup> schools. 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 Performance, Planning & Technology Branch that oversees programs promoting innovation and improving student achievement. Programs include oversight of statewide assessments and the collection and reporting of educational data (CDE, 2018d). Within the Performance, Planning & Technology Branch, the Assessment Development & Administration Division manages the development and administration for all statewide assessments.

## 1.7.3. 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 in regard to population types, were invited to participate in the entire CAA for Science assessment development process. These California educators participated in tasks that included defining the purpose and scope of the assessment, assessment design, item development, data review, and score reporting.

## 1.7.4. Contractors

### 1.7.4.1 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 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 *Directions for Administration*

<sup>&</sup>lt;sup>2</sup> Retrieved from the CDE Fingertip Facts on Education in California – *CalEdFacts* web page at <u>https://www.cde.ca.gov/ds/sd/cb/ceffingertipfacts.asp</u>

- 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

## **1.8. Systems Overview and Functionality**

#### **1.8.1. 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 embedded PTs, CAASPP user information, and access to the CAASPP Security and Test Administration Incident Reporting System form

TOMS receives student enrollment data and LEA/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."<sup>3</sup> 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 2017–18 CAASPP Online Test Administration Manual (CDE, 2018a).

#### 1.8.2. Online Reporting System (ORS)

LEAs use the ORS to view participation results from the CAASPP assessments. The primary purpose of the ORS is for LEAs to access completion data to determine which students need to complete testing or start testing.

## **1.9. Overview of the Technical Report**

This technical report addresses the characteristics of the CAAs for Science administered starting in November 2017 and contains nine additional chapters as follows:

<sup>&</sup>lt;sup>3</sup> From the CDE California Longitudinal Pupil Achievement Data System (CALPADS) web page at <u>http://www.cde.ca.gov/ds/sp/cl/</u>

- <u>Chapter 2</u> presents an overview of processes involved in the CAA for Science secondyear pilot, including descriptions of item development, test administration, and psychometric analyses.
- <u>Chapter 3</u> discusses the detailed procedures of embedded PT development for the CAA for Science second-year pilot.
- <u>Chapter 4</u> describes the details of administering the embedded PTs for the CAA for Science second-year pilot, as well as the procedures followed by ETS to ensure test security.
- <u>Chapter 5</u> summarizes the scoring approaches and type of scores that are reported for the CAA for Science second-year pilot.
- <u>Chapter 6</u> summarizes the statistical procedures and results for 2017–18, including classical item analyses, test completion rates and analyses, and differential item functioning analyses.
- <u>Chapter 7</u> describes the development and administration of the survey questionnaires for test examiners and the results of analyses on their responses.
- <u>Chapter 8</u> 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 2017–18 CAA for Science second-year pilot.
- <u>Chapter 9</u> discusses the various procedures used to ensure the quality of the CAA for Science second-year pilot.
- <u>Chapter 10</u> 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. Retrieved from [this link is no longer active].
- California Department of Education. (2017, October). *State Board of Education responsibilities.* Retrieved from <u>http://www.cde.ca.gov/be/ms/po/sberesponsibilities.asp</u>
- California Department of Education. (2018a). CAASPP online test administration manual, 2017–18 administration. Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].
- California Department of Education. (2018b). *California Alternate Assessment for Science blueprint*. Sacrament, CA: California Department of Education. Retrieved from <a href="https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx">https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx</a>
- California Department of Education. (2018c, March). *California Alternate Assessments.* Retrieved from <u>http://www.cde.ca.gov/ta/tg/ca/altassessment.asp</u>
- California Department of Education. (2018d, July). *Organization.* Retrieved from <u>http://www.cde.ca.gov/re/di/or/</u>
- Gong, B., & Marion, S. (2006). Dealing with flexibility in assessments for students with significant cognitive disabilities. *Synthesis Report, 60*. Retrieved from [this link is no longer active].

# Chapter 2: Overview of the CAA for Science Second-Year Pilot Processes

This chapter provides an overview of the processes implemented by Educational Testing Service (ETS) during the full testing cycle for the 2017–18 California Alternate Assessment (CAA) for Science, including test development and administration, score production, and reporting. In addition, test participation is also described.

## 2.1. Embedded Performance Task (PT) 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).

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

For the second-year pilot, ETS developed three embedded PTs for each grade band according to the draft blueprint. The State Board of Education–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 draft blueprint was developed in consultation with the California Department of Education (CDE). It consisted of a Science Connector prioritization plan based on input from California educators, other internal and external experts on the CA NGSS, and alternate assessments. Each of the embedded PTs assesses one of these Science Connectors.

## 2.1.2. 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. In order to allow for the widest possible range of student participation, 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 beginning sentences 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 that 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; for this review of embedded PTs for the CAA for Science, the diverse needs of students with the most significant cognitive disabilities are carefully considered.

# 2.1.3. Embedded PTs Developed 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 then associated with a particular concept or phenomenon. The concept or topic selected for each 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 <u>chapter 3</u>.

## 2.2. Test Administration

The CAA for Science second-year pilot content and materials were provided in electronic PDF format. Authorized school and local educational agency (LEA) staff downloaded the *Directions for Administration (DFA)* for each embedded PT from the secure Test Operations Management System. Test examiners used the *DFA* materials in printed or electronic format.

During administration, test examiners recorded points earned by the student on the provided Answer Recording Document. After testing, LEAs entered these results into the online Data Entry Interface (DEI) for secure transmittal to ETS for analysis. LEAs were instructed to retain Answer Recording Documents as a local record of student score results.

### 2.2.1. Accessibility Features

For the administration of the CAA for Science embedded PTs, teachers were required to offer the same instructional resources and classroom accommodation(s) to each student that are customarily provided in accordance with the student's individualized education program (IEP) or Section 504 plan. These instructional resources and accommodations also apply to the collection of student responses for the CAAs for Science. Because the CAA for Science second-year pilot was not administered using the online California Assessment of Student Performance and Progress (CAASPP) testing interface, the CAASPP embedded universal tools, designated supports, and accommodations did not apply.

### 2.2.2. 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 <u>4.5 Accessibility Features for the Second-Year Pilot</u>.

## 2.3. Participation

The decision to assign a student to take the CAA for Science is made by his or her IEP team, which uses the information on the CAA Guidance for IEP Teams web page to make that determination. This web page describes the CAA and its administration, criteria for participation, and the students who should be assigned to take this test (CDE, 2018).

A student must meet all three of the following criteria to participate in the CAA:

1. A student with 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 someone to live independently and to function safely in daily life.

- 2. The student is learning content derived from the California Common Core State Standards (CCSS) and CA NGSS. Goals and instruction listed in the IEP for the student are linked to the enrolled grade-level CCSS and address knowledge and skills that are appropriate and challenging for this student.
- 3. The student's need for extensive, direct individualized instruction and substantial supports to achieve measurable gains in the grade-level and age-appropriate curriculum. The student:
  - a. Requires extensive, repeated, individualized instruction and support that is not of a temporary or transient nature; and
  - b. Uses 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 are eligible to take the CAAs are required to participate. All students were administered the Student Response Check (SRC) for each of the three embedded PTs. Students who do not provide a consistent, observable response to the SRC are not required to be administered the rest of the embedded PT.

Refer to <u>appendix 2.A</u> for a summary of the number of participants and the percent of participation of all students and the selected student groups for each test during the 2017–18 administration. Because the data in the *Number of Participants* rows includes students who answered at least one item on each of the three embedded PTs, the number of participants does not include students for whom the test was ended as a result of the SRC.

# 2.4. Scores

Student responses to each test item were scored locally by test examiners based on the scoring rubrics in the embedded PT document. Locally assigned scores of student responses, information on any individualizations of the task for the student, and student survey responses were recorded locally on Answer Recording Documents and then entered into the DEI.

## 2.4.1. Score Reporting

There were no individual student scores reported for the 2017–18 CAA for Science secondyear pilot. The ETS psychometric team prepared an aggregate data file of students' percent correct scores and the associated preliminary indicator category for LEAs.

## 2.4.2. Aggregation Procedures

To provide meaningful results to the stakeholders, CAA for Science results for a given grade-level assessment are aggregated and generated at the school, LEA or direct funded charter school, county, and state levels. State-level results are available on the CAASPP Results website at [this link is no longer active].

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 <u>5.2.3 Aggregate Score Reporting</u>. Aggregated scores that summarize student performance by grade for selected groups of students are provided in <u>Table 2.A.2</u> through <u>Table 2.A.4</u>. 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 <u>Table 5.6</u>.

## 2.5. Overview of Psychometric Analyses

### 2.5.1. Analysis Types

The psychometric analyses for the CAA for Science second-year pilot consisted of classical item analyses and differential item functioning (DIF) 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 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 <u>6.2 Classical Item Analyses</u>; <u>appendix 6.A</u> contains the results of the classical analyses.

Additionally, responses to both the student survey and the test examiner survey were analyzed to evaluate how material choices and individualization might have impacted student performance. One unique aspect of the CAA for Science second-year pilot 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, despite ETS concerns about the potential impact of this flexibility. Refer to <u>Chapter 8: Embedded Performance Task and Test Comparability Considerations</u> for details of these additional analyses.

### 2.5.2. Description of DIF Analyses

DIF analyses are conducted to detect possible test bias by locating items on which one group of students performs significantly better than another group. DIF is a collection of statistical methods utilized to recognize if performance varies across different groups of students (e.g., Male versus Female or White versus Black or African American).

If an item performs differentially across student groups, even when students are 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 and determine the sources and meanings of performance differences. Refer to subsection <u>6.5. Differential Item Functioning (DIF)</u> for DIF analyses conducted, and <u>appendix 6.D</u> for DIF analysis results.

## References

California Department of Education. (2018). CAA guidance for IEP teams. Retrieved from http://www.cde.ca.gov/ta/tg/ca/caaiepteamrev.asp

Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service. Retrieved from <u>https://www.ets.org/s/about/</u> <u>pdf/standards.pdf</u>

## **Appendix 2.A Participation Rates**

**Note:** In <u>Table 2.A.1</u>, the number of enrolled students refers to the number of students selected to test.

#### Table 2.A.1 CAA for Science Second-Year Pilot Participation Rates—Enrolled Students

Variable	Grade 5	Grade 8	HSGrade 10	HSGrade 11	HSGrade 12	HS—All Grades
Number of Enrolled Students	5,766	5,807	186	1,789	8,259	10,234
Number of Participants	3,940	3,879	64	979	4,522	5,565
Percent of Participation	68%	67%	34%	55%	55%	54%

Crown	Number of Students	Number of Participants	Percent of Participation
Group All students	5,766	3,940	68%
Gender: Male	3,922	2,691	69%
Gender: Female	1,844	1,249	68%
Ethnicity: American Indian or Alaska Native	33	26	79%
Ethnicity: Asian	438	288	66%
Ethnicity: Native Hawaiian or Other Pacific Islander	28	18	64%
Ethnicity: Filipino	137	95	69%
Ethnicity: Hispanic or Latino	3,296	2,326	71%
Ethnicity: Black or African American	441	286	65%
Ethnicity: White	1,178	763	65%
Ethnicity: Two or more races	164	109	66%
English proficiency: English only	3,406	2,233	66%
English proficiency: Initially fluent English proficient	42	28	67%
English proficiency: English learner	1,947	1,389	71%
English proficiency: Reclassified fluent English proficient	362	286	79%
English proficiency: To be determined	4	0	0%
English proficiency: English proficiency unknown	5	4	80%
Economic status: Not economically disadvantaged	2,009	1,233	61%
Economic status: Economically disadvantaged	3,757	2,707	72%
Primary disability: Intellectual disability	2,014	1,452	72%
Primary disability: Hearing impairment	42	30	71%
Primary disability: Speech or language impairment	177	150	85%
Primary disability: Visual impairment	32	9	28%
Primary disability: Emotional disturbance	39	24	62%
Primary disability: Orthopedic impairment	284	125	44%
Primary disability: Other health impairment	296	223	75%
Primary disability: Specific learning disability	419	351	84%
Primary disability: Deaf-blindness	0	0	0%
Primary disability: Multiple disabilities	348	145	42%
Primary disability: Autism	2,088	1,417	68%
Primary disability: Traumatic brain injury	27	14	52%
Primary disability: Not classified	0	0	0%

# Table 2.A.2 CAA for Science Second-Year Pilot Participation Rates for Grade Five by Student Group

Group	Number of Students	Number of Participants	Percent of Participation
All students	5,807	3,879	67%
Gender: Male	3,904	2,599	67%
Gender: Female	1,903	1,280	67%
Ethnicity: American Indian or Alaska Native	34	22	65%
Ethnicity: Asian	491	311	63%
Ethnicity: Native Hawaiian or Other Pacific Islander	26	15	58%
Ethnicity: Filipino	175	122	70%
Ethnicity: Hispanic or Latino	3,201	2,250	70%
Ethnicity: Black or African American	478	300	63%
Ethnicity: White	1,227	750	61%
Ethnicity: Two or more races	128	78	61%
English proficiency: English only	3,383	2,176	64%
English proficiency: Initially fluent English proficient	97	70	72%
English proficiency: English learner	1,773	1,213	68%
English proficiency: Reclassified fluent English proficient	545	416	76%
English proficiency: To be determined	3	0	0%
English proficiency: English proficiency unknown	6	4	67%
Economic status: Not economically disadvantaged	2,142	1,242	58%
Economic status: Economically disadvantaged	3,665	2,637	72%
Primary disability: Intellectual disability	2,255	1,614	72%
Primary disability: Hearing impairment	55	31	56%
Primary disability: Speech or language impairment	112	82	73%
Primary disability: Visual impairment	37	13	35%
Primary disability: Emotional disturbance	47	26	55%
Primary disability: Orthopedic impairment	274	116	42%
Primary disability: Other health impairment	318	221	69%
Primary disability: Specific learning disability	364	274	75%
Primary disability: Deaf-blindness	6	0	0%
Primary disability: Multiple disabilities	341	134	39%
Primary disability: Autism	1,972	1,350	68%
Primary disability: Traumatic brain injury	26	18	69%
Primary disability: Not classified	0	0	0%

# Table 2.A.3 CAA for Science Second-Year Pilot Participation Rates for Grade Eight byStudent Group

	er of nts	Number of Participants	Percent of Participation
Group	Number of Students	Number of Participant	Perce Partic
All students	10,234	5,565	54%
Gender: Male	6,548	3,571	55%
Gender: Female	3,686	1,994	54%
Ethnicity: American Indian or Alaska Native	75	47	63%
Ethnicity: Asian	662	361	55%
Ethnicity: Native Hawaiian or Other Pacific Islander	45	23	51%
Ethnicity: Filipino	371	182	49%
Ethnicity: Hispanic or Latino	5,686	3,148	55%
Ethnicity: Black or African American	1,024	558	54%
Ethnicity: White	2,121	1,131	53%
Ethnicity: Two or more races	196	91	46%
English proficiency: English only	5,733	3,134	55%
English proficiency: Initially fluent English proficient	161	74	46%
English proficiency: English learner	3,231	1,714	53%
English proficiency: Reclassified fluent English proficient	1,100	637	58%
English proficiency: To be determined	2	1	50%
English proficiency: English proficiency unknown	7	5	71%
Economic status: Not economically disadvantaged	3,714	1,833	49%
Economic status: Economically disadvantaged	6,520	3,732	57%
Primary disability: Intellectual disability	4,363	2,609	60%
Primary disability: Hearing impairment	114	60	53%
Primary disability: Speech or language impairment	64	43	67%
Primary disability: Visual impairment	79	29	37%
Primary disability: Emotional disturbance	93	37	40%
Primary disability: Orthopedic impairment	833	280	34%
Primary disability: Other health impairment	362	196	54%
Primary disability: Specific learning disability	467	285	61%
Primary disability: Deaf-blindness	4	2	50%
Primary disability: Multiple disabilities	587	197	34%
Primary disability: Autism	3,184	1,784	56%
Primary disability: Traumatic brain injury	84	43	51%
Primary disability: Not classified	0	0	0%

# Table 2.A.4 CAA for Science Second-Year Pilot Participation Rates for High School by<br/>Student Group

# Chapter 3: Embedded Performance Task (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.

## 3.1. Embedded PT 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.

### 3.1.1. Specifications for the Embedded PTs

The item specifications for embedded PTs 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 performance expectation (PE)
- The full statement of the Science Connector
- The full content of each assessed FSKA 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

The *Directions for Administration (DFA)* for each CAA for Science PT administered during the second-year pilot contained the following:

- Task standards table
- Task materials list
- Student Response Check (refer to subsection <u>4.3.1 Administration of the Student</u> <u>Response Check</u>)
- Activities and their associated items

- Student survey (questions answered by the student and entered by the test examiner; refer to subsection <u>7.2 Student Survey Administration</u>)
- Student engagement survey (completed by test examiner after testing; refer to subsection <u>5.3.1 Student Engagement</u>)
- Answer Recording Document
- Related graphics and tables, if applicable

### 3.1.2. Embedded PT Format

Embedded PTs for the CAA for Science are designed to be engaging to the target population. Embedded PTs are developed with the understanding that a test examiner will 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* begins with background information and instructions for the test examiner. These instructions include the following:

- General steps for administering the assessment
- Modifications of materials and activities
- How to use the embedded PT, including specific directions for administering the assigned embedded PT, with details on the following features:
  - Composed of one or more items with one or more steps
  - Indicators on whether the step is an action for the test examiner to take or a question for the student to answer
  - Steps the test examiner must take after an item or collection of items to record the student's level of independence
  - Exemplar-required activity
  - Student Answer Recording Document
  - Answer key
  - CAA for Science student survey

The CAA for Science include the following item formats:

- **Selected Response**—Students are instructed to select one or more choices. Most CAA for Science items have two or three options; a few items have four options.
- **Identification**—Students are instructed to place the label on a specified part of a diagram or chart.

All scoring the selected-response and identification items is done by test examiners during the test administration. Scoring rubrics specific to each item are included in the *DFA* and are used by the test examiners to rate students' responses.

The number of items and points for each embedded PT is provided in <u>Table 3.1</u>.

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	7	9	6	6	6	6	19	21
Grade 8	6	8	4	8	3	11	13	27
High School	3	8	7	7	6	6	16	21

#### Table 3.1 Number of Items and Points for Each PT

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

#### 3.1.3.1 Recruitment

Applications for embedded PT item writing are screened by senior ETS content staff. Only those applicants with strong science content and teaching backgrounds are approved for inclusion in the training program for item writing. All embedded PT item-writing participants are current or former California educators who are particularly knowledgeable about the CA NGSS assessed by the CAA for Science and are experienced with the test-taking population.

All item writers meet 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 is desirable
- Current experience teaching science in one or more grades in grades five through twelve in California, and when possible, experience teaching students with cognitive disabilities
- 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 areas, or both
- Familiarity, understanding, and support of the Science Connectors, EUs, and FKSAs

### 3.1.4. Embedded PT Item Writer Training

Item writer training for the second-year pilot cycle took place over three days in July 2017. 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.

## 3.2. Embedded PT Review Process

### 3.2.1. Selection of Embedded PTs

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

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

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

The ETS review process for the CAA for Science includes 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 evaluate the activities and items in adherence to the rules for item development.

## 3.2.2. ETS Content Review

Embedded PTs undergo three rounds of content reviews by content-area assessment specialists with increasing levels of expertise, called Round 1, Round 2, and Final Round. These assessment specialists ensure that the embedded PTs are in compliance 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 tier 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 is classified with the Science Connector, EU, and the FKSA it is intended to measure. Assessment specialists check each item against its classification

codes, both to evaluate the correctness of the classification and to ensure that the task posed by the item is relevant to the outcome it was intended to measure. The reviewers can accept the item and classification as written, suggest revisions, or recommend that the item be discarded. These steps occur prior to the CDE's review.

#### 3.2.3. ETS Editorial Review

After content-area assessment specialists review each item, a group of specially trained editors also review each embedded PT and item in preparation for consideration by the CDE and California educators. The editors check 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.

### 3.2.4. 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—conduct the next level of review. These trained staff members review every item before the CDE and formal embedded PT item reviews.

The review process promotes 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

## 3.3. California Educator's Review

### 3.3.1. California Educators as Content Experts

Meetings with California educators are 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 fill an advisory role to the CDE and ETS and provide guidance on matters related to embedded PT item development for the CAA for Science. These educators are responsible for reviewing all newly developed items for alignment to the California content standards. Meeting participants also review the items for the accuracy of content, clarity of phrasing, and quality. In their examination of embedded PT items, participants can raise concerns related to age or grade appropriateness as well as gender, racial, ethnic, or socioeconomic bias.

### 3.3.2. Composition of Item Review Panels

California educators participating in item review meetings consist of current and former teachers, resource specialists, administrators, curricular experts, and other education professionals. Minimum qualifications to be invited to participate are

- three or more years of teaching experience in grades kindergarten through twelve,
- three or more years of teaching experience in science,

Embedded Performance Task (PT) Development and Review | California Educator's Review

- bachelor's or higher degree in a grade or content area related to science, and
- knowledge and experience with the CA NGSS.

Preferred qualifications include

- special education credential,
- 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 meet 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;
- Bachelor's or higher degree in science; and
- Knowledge of and experience with the CA NGSS.

Every effort is made to ensure that groups of item reviewers include both genders as well as a wide representation of geographic regions and ethnic groups in California. Efforts also are made to ensure representation by members with experience serving California's diverse special education population.

<u>Table 3.2</u> shows the educational qualifications, present occupation, and credentials of the individuals who participated in CAA for Science item review.

Qualification Type	Qualification	Number of Reviewers
NA	Total number of reviewers	7
Occupation	Teacher or Program Specialist, Elementary School	3
Occupation	Teacher or Program Specialist, Middle School	2
Occupation	Teacher or Program Specialist, High School	2
Occupation	Other District Personnel	0
Highest Degree Earned	Bachelor's Degree	2
Highest Degree Earned	Master's Degree	5
Highest Degree Earned	Doctorate	0

 Table 3.2 Number of Item Reviewers with Each Qualification

Qualification Type	Qualification	Number of Reviewers
K–12 Teaching Credential	Elementary Teaching (multiple subjects)	2
K–12 Teaching Credential	Secondary Teaching (single subject)	5
K–12 Teaching Credential	Special Education	4
K–12 Teaching Credential	Reading Specialist	0
K–12 Teaching Credential	English Learner (CLAD, BCLAD)	0
K–12 Teaching Credential	Administrative	0
K–12 Teaching Credential	Other	0

**Note:** Some reviewers have multiple occupations or teaching credentials. Some are currently working toward earning their highest degree.

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

## 3.3.3. Meetings for Review of CAA for Science Items

ETS science assessment specialists facilitate CAA Science item review meetings. Each meeting begins with a brief training session on how to review embedded PT items. ETS provides this training, which consists 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 include the following:

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

Criteria also encompass 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 are trained on how to make recommendations for revising items.

Guidelines for reviewing items are 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 test 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?

## 3.4. Data Review

After items have been included in an operational or field test and administered to students, ETS prepares the items and the associated statistics for review by the CDE and California educators. Review materials include embedded PT items with their statistical data along with annotated comment sheets for use by reviewers. ETS conducts an introductory training to highlight any new issues and serve as a statistical refresher. Reviewers then make decisions about which items should be included in the item bank for future assembly. If an item is considered problematic and not to be included in the item bank, it will be revised and once again follow the steps in the item development process, including field testing. ETS psychometric and content staff are available to reviewers throughout this process.

# **Chapter 4: Test Administration**

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

# 4.1. Grade Assignment for High Schools

All local educational agencies (LEAs) with eligible students in grades five, eight, and twelve administered the second-year pilot for CAA for Science. Students in grades ten and eleven 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. High school students in an ungraded program whose calculated grade was twelve took this assessment, as did students in grades ten or eleven, if assigned. A student's grade was calculated by subtracting five from his or her chronological age on September 1, 2017.

Students in grades five and eight and high school (grade ten, eleven or twelve) who met 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 content standards; and
- the student requires extensive direct individualized instruction and substantial resources to achieve measurable gains in the grade and age-appropriate curriculum.

## 4.2. Administration Preparations

The embedded PTs were designed to be administered to students in conjunction with the normal course of instruction related to the California Next Generation Science Standard Core Content Connector (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.

### 4.2.1. 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 Assessment Videos and Archived Webcasts web page at <u>http://www.caaspp.org/training/caaspp/</u>. For the CAA for Science, two videos were produced.

The first video, [this link is no longer active] included the following topics and demonstrations:

- Overview of 2017–18 CAA for Science second-year pilot
- · 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

The second video, [this link is no longer active] provided a step-by-step guide on entering student results into the Data Entry Interface (DEI).

Finally, ETS produces an online module, the CAA Test Examiner Tutorial, designed to teach test examiners on how to administer the CAAs, including 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. This video is available on the CAAs web page at <a href="http://www.caaspp.org/administration/about/caa/">http://www.caaspp.org/administration/about/caa/</a>.

### 4.2.2. Practice and Training Tests

Sample embedded PTs were available for training at all assessed grade levels (California Department of Education [CDE], 2018a).

## 4.3. Test Administration

The CAA for Science second-year pilot 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 available as electronic PDFs that the test examiner had the option of printing. Results were recorded by the test examiner on the provided Answer Recording Document.

LEAs were asked to retain completed Answer Recording Documents and other paper materials that were used for transcribing student responses into the DEI. Results entered in the DEI included student score points earned on each test question, survey responses, and individualizations.

### 4.3.1. Administration of the Student Response Check (SRC)

Prior to beginning the embedded performance task (PT), the test examiner conducted an 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 second-year pilot, 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 PT. The test examiner showed the objects to the student and directed him or her 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.

### 4.3.2. 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.

#### 4.3.2.1 Actions to Administer

Test examiners followed these steps in administering an embedded PT to a student:

- 1. Decide on an administration date in line with the normal course of instruction
- 2. Review the activities associated with the embedded PT; based on this review of activity materials, setup, and processes, determine whether the exemplar activity or an individualized activity will be presented
  - If an individualized activity was presented, direction scripts were also individualized.
  - If a task contains an activity that is not easily repeatable in multiple one-on-one administrations, a test examiner may present that activity in a group setting or use a video of that activity to present to students. The video can then be used in the one-on-one administration of the PT.
- 3. Gather and print necessary activity materials
- 4. Print an Answer Recording Document for each student for recording results that will later be transcribed into the online DEI
- 5. Complete the preparation instructions, including printing graphics and any necessary individualization of the exemplar script
- 6. Administer the activity according to the administration instructions, verbalizing statements or performing indicated actions

#### 4.3.2.2 Actions to Record

Test examiners recorded task administration data for each student on the task Answer Recording Documents. After the embedded PT was administered, and once the DEI was made available, the test examiner or a CAASPP coordinator entered embedded PT results into the DEI.

#### 4.3.3. Administration of the Survey

After an embedded PT was administered to a student, the test examiner asked the student two questions as part of the student survey and then answered three questions about the level of student engagement. The results of these surveys were gathered along with student responses in the DEI.

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 <u>7.4 Test</u> <u>Examiner Survey Results</u> for more information about the feedback received.

## 4.4. Procedures to Maintain Standardization

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

#### 4.4.1. LEA CAASPP Coordinator

An LEA CAASPP coordinator was designated by the district superintendent at the beginning of the 2017–18 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 the STAIRS response email indicates that an appeal is warranted).

### 4.4.2. 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 test site coordinator is responsible for identifying test examiners and ensuring that they have signed CAASPP Test Security Affidavits. 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 participate, 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 the STAIRS response email indicates that an appeal is warranted).

#### 4.4.3. 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 for the CAA for Science*.

#### 4.4.4. Instructions for Test Examiners and Staff Involved in CAA Administration

#### 4.4.4.1 Directions for Administration (DFA)

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, 2018a).

#### 4.4.4.2 CAASPP Online Test Administration Manual

The CAASPP Online Test Administration Manual (CDE, 2018b) 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 include 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 include 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.

#### 4.4.4.3 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 access TOMS to retrieve *CAA for Science DFAs*.

TOMS modules include the following (CDE, 2018e):

- **Test Administration Setup**—This module allows LEAs to determine and calculate dates for the LEA's 2017–18 testing.
- Adding and Managing Users—This module allows LEA CAASPP coordinators to add CAASPP test site coordinators and test examiners to TOMS so that the designated user can access the online embedded PT *DFAs*.
- **Student Test Assignment**—This module allows LEA CAASPP coordinators to designate students to take the alternate assessments.

#### 4.4.4.4 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, 2018d).
- Security Incidents and Appeals Procedure Guide—This manual provides information on how to report and submit an appeal to the CDE to reset, reopen, invalidate, or restore individual online student assessments (CDE, 2018c).

## 4.5. Accessibility Features for the Second-Year Pilot

#### 4.5.1. Individualizations

A notable feature of the 2017–18 embedded PTs is that test examiners had the option to individualize certain elements of the assessment, although not all PTs allowed for individualization. For the second-year pilot 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, including revisions to administration scripts and material choices, on the Answer Recording Document, which were later transcribed into the DEI.

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

<u>Table 4.1</u> through <u>Table 4.3</u> display the results of the test examiner survey regarding the kinds of individualization provided. N-counts in these tables are based on all students in version 4 of the production file ("P4") released on October 9, 2018, with an include indicator of "T" to indicate the student tested.

Individualization	PT 1 (Sun and Shadows)— Number	PT 1 (Sun and Shadows)— Percent of Total	PT 2 (Physical Changes)— Number	PT 2 (Physical Changes)— Percent of Total	PT 3 (Weather Conditions)— Number	PT 3 (Weather Conditions)— Percent of Total
Using Standardized Scripts	4,472	97%	4,356	97%	4,770	98%
Using Individualized Scripts	147	3%	141	3%	89	2%
Using Standardized Diagram	4,471	96%	NA	NA	NA	NA
Using Individualized Diagram	204	4%	NA	NA	NA	NA
Using Standardized Materials	NA	NA	3,542	79%	NA	NA
Using Individualized Materials	NA	NA	969	21%	NA	NA

 Table 4.1 Individualizations—Grade Five

		4.2 mai										
Individualization	PT 1 (Water Cycle), Activity 1— Number	PT 1 (Water Cycle), Activity 1— Percent of Total	PT 1 (Water Cycle), Activity 2— Number	PT 1 (Water Cycle), Activity 2— Percent of Total	PT 2 (Bioenergy), Activity 1— Number	PT 2 (Bioenergy), Activity 1— Percent of Total	PT 2 (Bioenergy), Activity 2— Number	PT 2 (Bioenergy), Activity 2— Percent of Total	PT 3 (Cells), Activity 1—Number	PT 3 (Cells), Activity 1—Percent of Total	PT 3 (Cells), Activity 2—Number	PT 3 (Cells), Activity 2—Percent of Total
Using Standardized Scripts	4,350	98%	4,337	98%	NA	NA	4,318	98%	4,299	97%	4,309	98%
Using Individualized Scripts	94	2%	83	2%	NA	NA	82	2%	111	3%	85	2%
Using Standardized Diagram or Picture	3,990	89%	4,044	91%	3,937	89%	NA	NA	NA	NA	NA	NA
Using Individualized Diagram or Picture	472	11%	397	9%	474	11%	NA	NA	NA	NA	NA	NA
Using Standardized Materials	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Using Individualized Materials	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

#### Table 4.2 Individualizations—Grade Eight

	PT 1 (Molecules), Activity 1— Number	PT 1 (Molecules), Activity 1— Percent of Total	PT 1 (Molecules), Activity 2— Number	PT 1 (Molecules), Activity 2— Percent of Total	PT 2 (Force and Motion), Activity 1—Number	PT 2 (Force and Motion), Activity 1—Percent of Total	PT 2 (Force and Motion), Activity 2—Number	PT 2 (Force and Motion), Activity 2—Percent of Total	PT 3 (Erosion), Activity 1— Number	PT 3 (Erosion), Activity 1— Percent of Total
Individualization					-	- 1				
Using Standardized Scripts	6,431	96%	6,445	97%	6,366	97%	NA	NA	6,339	97%
Using Individualized Scripts	237	4%	190	3%	171	3%	NA	NA	175	3%
Using Standardized Diagram	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Using Individualized Diagram	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Using Standardized Materials	NA	NA	NA	NA	5,371	82%	5,283	81%	5,194	81%
Using Individualized Materials	NA	NA	NA	NA	1,207	18%	1,248	19%	1,248	19%

 Table 4.3 Individualizations—High School

### 4.5.2. 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 that otherwise may not engage at all with the activity or item.

### 4.5.3. Choice of Materials

The activities that are part of each embedded PT are almost all designed to allow test examiners to substitute different materials, as long as the required activity is administered. Test examiners are 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 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).

## 4.5.4. 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 that are customarily provided in accordance with the student's IEP. These instructional supports and accommodations also apply to the collection of student responses for the CAA for Science.

## 4.6. Test Security and Confidentiality

## 4.6.1. ETS' Office of Testing Integrity (OTI)

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

#### 4.6.2. 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, 2018a).

The following measures ensured the security of CAASPP System assessments administered in 2017–18:

- 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 to the California Technical Assistance Center before ETS granted the coordinators access to TOMS. (5 *CCR*, Education, Section 859[a])
- Anyone having access to the testing materials must have 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 to the CAASPP test site coordinator 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])

#### 4.6.3. 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.

Additionally, 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.

#### 4.6.4. 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.

### 4.6.5. 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.

### 4.6.6. Statistical Analysis on Secure Servers

During CAASPP testing, ETS information technology staff retrieves data files from the American Institutes for Research and loads those files into a database. 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.

### 4.6.7. 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.

### 4.6.8. Student Test Results

#### 4.6.8.1 Types of Results

No individual Student Score Reports were provided for the CAA for Science second-year pilot. During testing, test examiners used Answer Recording Documents provided inside each embedded PT *DFA* to record student results. After results were entered into the DEI, LEAs were directed to keep the Answer Recording Documents as a record of student results. Additionally, LEAs were provided with tools and guidelines for deriving preliminary indicators based on the results from the Answer Recording Document. These are as follows:

- <u>Preliminary Indicator Conversion Table for the CAA for Science</u>—This document provides step-by-step instructions to calculate preliminary indicators (i.e., percent correct and indicator category).
- Preliminary Indicator Calculator for the CAA for Science [this link is no longer active].
   —This optional tool automatically calculates the preliminary indicator category for a student who took the 2017–18 CAA for Science. Results for up to 10 students at a time could be calculated using this Excel spreadsheet.
- Internet reports—Requestable reports of student results are aggregated by state, county, LEA, and test site.

The preliminary indicator and the aggregate report information are described in <u>chapter 5</u>.

#### 4.6.8.2 Security of Results Files

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

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

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

Access to the ETS Computer Processing Center is controlled by employee and visitor identification badges. The Center is secured by doors that only can be unlocked by the badges of personnel who have functional responsibilities within its secure perimeter. Authorized personnel accompany visitors to the ETS Computer Processing Center at all times. Extensive smoke detection and alarm systems, as well as a pre-action fire-control system, are installed in the Center.

#### 4.6.8.3 Security of Individual Results

ETS protects individual students' results during the following events:

- Scoring
- Transfer of scores by means of secure data exchange
- Reporting
- Posting of aggregate data
- Storage

In addition to protecting the confidentiality of testing materials, the ETS Code of Ethics further prohibits ETS employees from financial misuse, conflicts of interest, and unauthorized appropriation of ETS property and resources. Specific rules are also given to

ETS employees and their immediate families who may take a test developed by ETS (e.g., the CAA for Science). The ETS Office of Testing Integrity (OTI) verifies that these standards are followed throughout ETS. This verification is conducted, in part, by periodic on-site security audits of departments, with follow-up reports containing recommendations for improvement.

#### 4.6.9. 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, 2018c). 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 filling out the secure STAIRS form for reporting, which contains selectable options to guide coordinators in their submittal. After the form is submitted, an email containing a case number and next steps will be sent to the submitter (and to the LEA CAASPP coordinator, if the form is submitted by the CAASPP test site coordinator). The CAASPP STAIRS form provides the LEA CAASPP coordinator, the CDE, and the California Technical Assistance Center (CaITAC) with the opportunity to interact and communicate regarding the STAIRS process. (CDE, 2018d)

The following types of STAIRS reports, as applicable to the CAA, are also forwarded to the CDE:

- Security breach (where secure materials are exposed)
- Accidental access to a summative assessment
- Incorrect Statewide Student Identifier used (intentionally switched)

#### 4.6.9.1 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 reported the incident within 24 hours, using the online *CAASPP STAIRS* form.

#### 4.6.9.2 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. An irregularity must be reported to the LEA CAASPP coordinator and CAASPP test site coordinator immediately. The coordinator reported the irregularity within 24 hours, using the online *CAASPP STAIRS* form.

#### 4.6.9.3 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 complete the online *CAASPP STAIRS* form 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.

## References

- California Department of Education. (2018a). CAA for Science 2018 Sample Task: Grade 5 embedded performance task, "Fossils." Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].
- California Department of Education. (2018b). CAASPP online test administration manual, 2017–18 administration. Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].
- California Department of Education. (2018c). Security incidents and appeals procedure guide, 2017–18 administration. Sacramento, CA: California Department of Education.
- California Department of Education. (2018d). *Technical specifications and configuration guide for CAASPP online testing*. Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].
- California Department of Education. (2018e). *TOMS pre-administration guide for CAASPP testing*. Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].
- Educational Testing Service. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service. Retrieved from <u>https://www.ets.org/s/about/pdf/</u> <u>standards.pdf</u>

# **Chapter 5: Scoring and Reporting**

Student scores for the second-year pilot of the California Alternate Assessment (CAA) for Science, given during the 2017–18 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.

## **5.1. CAA for Science Scoring Process**

Each student was administered three embedded performance tasks (PTs), each consisting of three to seven items. During the test administration, the test examiner scored the student's responses for each item and then recorded the student's scores on the Answer Recording Document provided for each embedded PT. The test examiner or LEA staff entered these results into the online Data Entry Interface (DEI). Instructions detailing how to score the student responses and how to enter the results into the DEI were provided by Educational Testing Service (ETS) in the *Embedded Performance Task Directions for Administration* (California Department of Education [CDE], 2018a).

The test examiner or LEA staff then computed the overall percent correct scores (described in subsection <u>5.2.1 Percent Correct</u>) and preliminary indicator category (described in subsection <u>5.2.2 Preliminary Indicator Categories</u>) for each student. ETS provided a scoring tool to help the LEAs calculate the student's overall score, percent-correct score, and preliminary indicator category. More information about the preliminary indicators can be found on the CDE [this link is no longer active]. web page (CDE, 2018b).

Although there was no formal Student Score Report for the CAA for Science second-year pilot, 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.

## 5.2. Types of Scores

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

## 5.2.1. Percent Correct

The test examiner recorded the points the student earned on each item (e.g., a 0, 1, or 2 for a 2-point item) on an Answer Recording Document. Using the three completed Answer Recording Documents, either the test examiner or LEA staff added the results of the three scores and then calculated the overall percent-correct score, which is the total score divided by the total maximum number of points possible. When the student did not respond to at least one item for the PT, a score of 0 (zero) was assigned for that PT.

### 5.2.2. 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, 2018c).

There are three preliminary indicator categories to indicate high (category 3), medium (category 2), or low (category 1) performance. A student's preliminary indicator category provides a general indication of the student's understanding of the Science Connectors. <u>Table 5.1</u> provides the description of each indicator category.

Category	Explanation
3	Student performance suggests a <i>considerable</i> understanding of the Science Connectors.
2	Student performance suggests a <i>moderate</i> understanding of the Science Connectors.
1	Student performance suggests a <i>limited</i> understanding of the Science Connectors.

#### Table 5.1 Indicator Categories

Students who performed at or below the chance level—the average performance expected of students responding to each item at random—are assigned to the indicator category of 1. Students who performed exceedingly well (i.e., 90 percent correct or above) are 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 cut scores for the three indicator categories are shown in <u>Table 5.2</u>. Each cut score is expressed as a percentage of the maximum possible score.

#### Table 5.2 Cut Scores for Preliminary Categories

Grade Level	Required for Category 2	Required for Category 3
Grade 5	34%	90%
Grade 8	32%	90%
High School	29%	90%

Preliminary indicator conversion tables are shown for grades five (<u>Table 5.3</u>) and eight (<u>Table 5.4</u>) and high school (<u>Table 5.5</u>). These tables provide the percent-correct score and preliminary category for each possible raw score.

Raw Score	Deveent	Drolinsinom
(# of points	Percent	Preliminary
earned)	Correct	Category
0	0.0	1
1	4.8	1
2	9.5	1
3	14.3	1
4	19.0	1
5	23.8	1
6	28.6	1
7	33.3	1
8	38.1	2
9	42.9	2
10	47.6	2
11	52.4	2
12	57.1	2
13	61.9	2
14	66.7	2
15	71.4	2
16	76.2	2
17	81.0	2
18	85.7	2
19	90.5	3
20	95.2	3
21	100.0	3

 Table 5.3 Grade Five Preliminary Indicator Conversion Table

 Table 5.4 Grade Eight Preliminary Indicator Conversion Table

Raw Score (# of points earned)	Percent Correct	Preliminary Category
0	0.0	1
1	3.7	1
2	7.4	1
3	11.1	1
4	14.8	1
5	18.5	1
6	22.2	1
7	25.9	1
8	29.6	1
9	33.3	2
10	37.0	2

Raw Score (# of points earned)	Percent Correct	Preliminary Category
11	40.7	2
12	44.4	2
13	48.1	2
14	51.9	2
15	55.6	2
16	59.3	2
17	63.0	2
18	66.7	2
19	70.4	2
20	74.1	2
21	77.8	2
22	81.5	2
23	85.2	2
24	88.9	2
25	92.6	3
26	96.3	3
27	100.0	3

Raw Score (# of points earned)	Percent Correct	Preliminary Category
0	0.0	1
1	4.8	1
2	9.5	1
3	14.3	1
4	19.0	1
5	23.8	1
6	28.6	1
7	33.3	2
8	38.1	2
9	42.9	2
10	47.6	2
11	52.4	2
12	57.1	2
13	61.9	2
14	66.7	2
15	71.4	2
16	76.2	2

Raw Score (# of points earned)	Percent Correct	Preliminary Category
17	81.0	2
18	85.7	2
19	90.5	3
20	95.2	3
21	100.0	3

<u>Table 5.A.1</u> 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.). <u>Table 5.A.2</u> and <u>Table 5.A.3</u> show the same information for the grade eight test and the high school test.

#### 5.2.3. 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 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 <u>appendix 5.A.</u> In <u>Table 5.A.1</u> through <u>Table 5.A.3</u>, 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.

<u>Table 5.6</u> 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 "NA."

Category	Student Groups
Gender	Male
	Female
Ethnicity	American Indian or Alaska Native
	• Asian
	<ul> <li>Native Hawaiian or Other Pacific Islander</li> </ul>
	• Filipino
	Hispanic or Latino
	<ul> <li>Black or African American</li> </ul>
	White
	Two or more races

 Table 5.6 Demographic Student Groups to Be Reported

Category	Student Groups
English-Language	English only
Fluency	<ul> <li>Initially fluent English proficient</li> </ul>
	<ul> <li>English learner</li> </ul>
	<ul> <li>Reclassified fluent English proficient</li> </ul>
	<ul> <li>To be determined</li> </ul>
	<ul> <li>English proficiency Unknown</li> </ul>
Economic Status	<ul> <li>Not economically disadvantaged</li> </ul>
	<ul> <li>Economically disadvantaged</li> </ul>
Primary Disability	Intellectual disability
Туре	<ul> <li>Hearing Impairment</li> </ul>
	<ul> <li>Speech or language impairment</li> </ul>
	<ul> <li>Visual Impairment</li> </ul>
	<ul> <li>Emotional disturbance</li> </ul>
	<ul> <li>Orthopedic impairment</li> </ul>
	<ul> <li>Other health impairment</li> </ul>
	<ul> <li>Specific learning disability</li> </ul>
	Deaf-blindness
	<ul> <li>Multiple disabilities</li> </ul>
	Autism
	<ul> <li>Traumatic brain injury</li> </ul>
	Not classified <sup>4</sup>
Migrant Status	<ul> <li>Eligible for the Title I Part C Migrant Program (Migrant)</li> </ul>
	<ul> <li>Not eligible for the Title I Part C Migrant Program (Nonmigrant)</li> </ul>

## 5.3. Survey Questions Regarding Test Administration

Four student survey questions were presented at the end of each embedded PT. The test examiner entered responses to these questions on the Answer Recording Document; these were subsequently entered into the DEI by the test examiner or LEA staff.

### 5.3.1. Student Engagement

In the survey that follows, the first two questions were presented to the student; the test examiner recorded the responses. The test examiner answered the final two questions regarding the student's mode of communication and level of engagement during administration of the embedded PT.

- 1) How did the student feel about taking this performance task?
  - A Happy
  - B Sad
  - $\circ$  C Confused
  - D No Response

<sup>&</sup>lt;sup>4</sup> Disability information was changed or removed after student testing.

- 2) Did the student have enough time to complete this performance task?
  - A Yes
  - **B No**
  - C No Response
- Select the mode(s) of communication used by the student on this performance task. (Select all that apply.)
  - □ Mouse, touchscreen, and/or a computer keyboard
  - □ Verbal response
  - □ Gestures or pointing
  - □ Written response
  - □ Assistive/augmentative communication device
  - Eye gaze
  - □ Other

If the student used a mode of communication that is not listed, please indicate it below. (Space was provided for the test examiner to provide a written response.)

- 4) 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 7.3 Student Survey Results.

#### 5.3.2. 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 second-year pilot 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, including revisions to administration scripts, on the Answer Recording Document, which were later transcribed into the DEI.

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

- 1) Did you use an alternative diagram?
  - A Yes

```
• B – No
```

- 2) Did you use an Individualized Script?
  - A Yes
  - **B No**

If you delivered an Individualized Script, please enter it below. If you did not deliver an Individualized Script, please skip. (Space was provided for the test examiner to provide a written response.)

Scoring and Reporting | Survey Questions Regarding Test Administration

- 3) Did you administer the Exemplar Activity or an Individualized Activity?
  - A Yes
  - **B No**
- 4) If you administered an Individualized Activity, please list the materials used. If you did not administer an Individualized Activity, please skip. (Select all that apply)
  - The materials listed varied depending upon the hands-on activity for the embedded PT.

The summary of the individualization is provided in subsection <u>4.5 Accessibility Features for</u> <u>the Second-Year Pilot</u>.

## References

California Department of Education. (2018a). CAA for Science 2018 Sample Task: Grade 5 embedded performance task, "Fossils." Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].

California Department of Education. (2018b). *Preliminary indicator communication toolkit.* 

Retrieved from [this link is no longer active].

California Department of Education. (2018c). *Science assessments preliminary Indicators* 

FAQ, question 7. Retrieved from [this link is no longer active].

## Appendix 5.A: Demographic Summaries

#### Notes:

- To protect privacy when the number of students in a student group is 10 or fewer, the summary statistics at the test and reporting levels are not reported and are presented as "NA" in the tables in this appendix.
- Percentages in these tables may not sum up to 100 due to rounding.

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
All Valid Scores	4,712	10.46	6.17	49.82	29.4	30%	61%	9%
Male	3,211	10.65	6.26	50.7	29.83	30%	60%	10%
Female	1,501	10.07	5.96	47.96	28.39	31%	63%	6%
American Indian or Alaska Native	29	13.66	6.22	65.02	29.64	21%	52%	28%
Asian	358	9.03	6	42.99	28.56	37%	59%	4%
Native Hawaiian or Other Pacific Islander	23	9.57	6.89	45.55	32.83	39%	43%	17%
Filipino	116	9.22	5.67	43.92	26.99	34%	61%	4%
Hispanic or Latino	2,765	10.62	6.14	50.56	29.23	29%	62%	9%
Black or African American	341	10.76	6.16	51.25	29.33	29%	60%	11%
White	901	10.63	6.22	50.63	29.6	29%	61%	10%
Two or more races	136	10.4	6.45	49.54	30.7	29%	62%	10%
English only	2,721	10.3	6.28	49.06	29.9	31%	60%	9%
Initially fluent English proficient	36	8.89	6.27	42.33	29.85	44%	50%	6%
English learner	1,626	10.59	6.03	50.45	28.74	28%	63%	9%

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
Reclassified fluent English proficient	322	11.38	5.84	54.19	27.82	25%	67%	8%
To be determined	2	NA	NA	NA	NA	NA	NA	NA
English proficiency unknown	5	NA	NA	NA	NA	NA	NA	NA
Not economically disadvantaged	1,545	9.44	6.25	44.94	29.78	37%	56%	7%
Economically disadvantaged	3,167	10.96	6.07	52.21	28.92	27%	64%	9%
Migrant	25	12.16	6.03	57.9	28.73	20%	68%	12%
Nonmigrant	4,687	10.45	6.17	49.78	29.4	30%	61%	9%
Intellectual disability	1,703	9.8	5.67	46.65	26.98	32%	63%	5%
Hearing impairment	32	12.16	5.82	57.89	27.72	25%	63%	13%
Speech or language impairment	150	14.51	3.67	69.08	17.47	2%	80%	18%
Visual impairment	17	5.53	7.56	26.33	36	65%	24%	12%
Emotional disturbance	27	13.93	4.21	66.31	20.07	0%	85%	15%
Orthopedic impairment	209	7.2	6.74	34.29	32.09	52%	44%	4%
Other health impairment	243	12.82	5.64	61.04	26.85	16%	71%	13%
Specific learning disability	352	16.54	3.15	78.77	15.02	1%	69%	30%
Deaf-blindness	0	NA	NA	NA	NA	NA	NA	NA
Multiple disabilities	266	5.38	5.7	25.64	27.14	65%	34%	1%
Autism	1,697	10.32	6.11	49.15	29.08	31%	60%	9%
Traumatic brain injury	16	11.94	5.57	56.85	26.51	13%	81%	6%
Not classified	0	NA	NA	NA	NA	NA	NA	NA

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged)	8	NA	NA	NA	NA	NA	NA	NA
Asian (Primary ethnicity—Not economically disadvantaged)	212	8.73	5.95	41.58	28.34	42%	55%	3%
Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged)	1	NA	NA	NA	NA	NA	NA	NA
Filipino (Primary ethnicity—Not economically disadvantaged)	76	9.41	5.64	44.8	26.84	36%	62%	3%
Hispanic or Latino (Primary ethnicity—Not economically disadvantaged)	503	9.09	6.47	43.26	30.81	40%	52%	8%
Black or African American (Primary ethnicity—Not economically disadvantaged)	97	10.04	6.12	47.82	29.14	32%	58%	10%
White (Primary ethnicity—Not economically disadvantaged)	537	10.06	6.23	47.9	29.66	32%	59%	9%
Two or more races (Primary ethnicity—Not economically disadvantaged)	89	10.03	6.06	47.78	28.85	30%	64%	6%
American Indian or Alaska Native (Primary ethnicity— Economically disadvantaged)	21	15.57	5.31	74.15	25.27	10%	52%	38%
Asian (Primary ethnicity—Economically disadvantaged)	146	9.46	6.06	45.04	28.85	32%	64%	5%
Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged)	22	9.59	7.06	45.67	33.6	41%	41%	18%
Filipino (Primary ethnicity—Economically disadvantaged)	40	8.88	5.78	42.26	27.54	33%	60%	8%

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
Hispanic or Latino (Primary ethnicity—Economically disadvantaged)	2,262	10.96	6.01	52.18	28.63	27%	64%	9%
Black or African American (Primary ethnicity— Economically disadvantaged)	244	11.05	6.16	52.62	29.35	28%	61%	11%

364 11.48

47 11.11

White (Primary ethnicity—Economically disadvantaged)

Two or more races (Primary ethnicity—Economically

disadvantaged)

6.11 54.64 29.09 25% 63%

7.14 52.89 34.01 26% 57% 17%

11%

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
All Valid Scores	4,512	15.14	8.51	56.09	31.51	22%	63%	14%
Male	3,013	15.39	8.53	56.99	31.59	22%	62%	16%
Female	1,499	14.65	8.44	54.27	31.27	23%	65%	11%
American Indian or Alaska Native	24	17.75	8.54	65.74	31.62	13%	50%	38%
Asian	368	13.36	8.07	49.48	29.87	27%	65%	8%
Native Hawaiian or Other Pacific Islander	21	12.62	9.13	46.74	33.81	38%	62%	0%
Filipino	142	14.74	7.81	54.59	28.91	20%	68%	12%
Hispanic or Latino	2,611	15.4	8.57	57.04	31.72	22%	63%	15%
Black or African American	345	15.33	8.43	56.77	31.22	21%	66%	13%
White	874	15.13	8.54	56.02	31.65	23%	62%	15%
Two or more races	90	15.2	8.65	56.3	32.03	23%	62%	14%
English only	2,541	15.16	8.54	56.13	31.65	22%	63%	15%
Initially fluent English proficient	87	12.4	8.25	45.93	30.56	29%	66%	6%
English learner	1,421	14.92	8.6	55.24	31.85	24%	62%	13%
Reclassified fluent English proficient	458	16.32	7.87	60.43	29.13	16%	67%	17%
To be determined	0	NA	NA	NA	NA	NA	NA	NA
English proficiency unknown	5	NA	NA	NA	NA	NA	NA	NA
Not economically disadvantaged	1,521	13.98	8.68	51.79	32.16	27%	62%	11%
Economically disadvantaged	2,991	15.73	8.36	58.27	30.95	20%	64%	16%

### Table 5.A.2 Demographic Summary for Grade Eight

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
Migrant	28	19.46	7.47	72.09	27.65	7%	64%	29%
Nonmigrant	4,484	15.12	8.51	55.99	31.51	23%	63%	14%
Intellectual disability	1,817	14.86	7.93	55.04	29.35	21%	69%	10%
Hearing impairment	36	17.39	8.45	64.4	31.31	14%	67%	19%
Speech or language impairment	82	20.63	6.37	76.42	23.58	6%	56%	38%
Visual impairment	22	9.41	9.55	34.85	35.37	55%	36%	9%
Emotional disturbance	28	22.32	6.19	82.67	22.92	4%	50%	46%
Orthopedic impairment	176	9.88	9.18	36.57	34	48%	47%	6%
Other health impairment	236	18.53	7.53	68.63	27.87	12%	65%	23%
Specific learning disability	274	23.34	3.85	86.44	14.25	0%	51%	49%
Deaf-blindness	2	NA	NA	NA	NA	NA	NA	NA
Multiple disabilities	254	7.69	8.86	28.46	32.83	58%	37%	4%
Autism	1,562	14.96	8.16	55.4	30.23	22%	65%	13%
Traumatic brain injury	23	15.26	9.92	56.52	36.75	26%	57%	17%
Not classified	0	NA	NA	NA	NA	NA	NA	NA
American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged)	6	NA	NA	NA	NA	NA	NA	NA
Asian (Primary ethnicity—Not economically disadvantaged)	197	12.45	8.39	46.12	31.09	31%	61%	7%
Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged)	7	NA	NA	NA	NA	NA	NA	NA

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
Filipino (Primary ethnicity—Not economically disadvantaged)	81	14.1	7.72	52.22	28.6	23%	67%	10%
Hispanic or Latino (Primary ethnicity—Not economically disadvantaged)	533	14.24	8.94	52.73	33.11	27%	61%	12%
Black or African American (Primary ethnicity—Not economically disadvantaged)	106	13.91	9.17	51.5	33.97	28%	61%	10%
White (Primary ethnicity—Not economically disadvantaged)	529	14.23	8.53	52.72	31.58	25%	63%	12%
Two or more races (Primary ethnicity—Not economically disadvantaged)	50	15.82	8.01	58.59	29.66	18%	68%	14%
American Indian or Alaska Native (Primary ethnicity— Economically disadvantaged)	18	17.61	8.68	65.23	32.17	11%	50%	39%
Asian (Primary ethnicity—Economically disadvantaged)	171	14.4	7.56	53.35	28	22%	69%	9%
Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged)	14	16.21	8.34	60.05	30.89	21%	79%	0%
Filipino (Primary ethnicity—Economically disadvantaged)	61	15.59	7.9	57.74	29.27	16%	69%	15%
Hispanic or Latino (Primary ethnicity—Economically disadvantaged)	2,078	15.7	8.44	58.15	31.27	21%	63%	16%
Black or African American (Primary ethnicity— Economically disadvantaged)	239	15.96	8.02	59.1	29.69	18%	69%	14%
White (Primary ethnicity—Economically disadvantaged)	345	16.49	8.4	61.08	31.12	19%	61%	20%
Two or more races (Primary ethnicity—Economically disadvantaged)	40	14.43	9.44	53.43	34.94	30%	55%	15%

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
All Valid Scores	6,956	11.4	7.33	54.26	34.9	28%	52%	21%
Male	4,434	11.64	7.36	55.45	35.02	27%	51%	22%
Female	2,522	10.96	7.26	52.18	34.58	30%	52%	18%
American Indian or Alaska Native	58	12.14	7.64	57.8	36.39	28%	43%	29%
Asian	451	10.69	7.09	50.89	33.78	29%	56%	15%
Native Hawaiian or Other Pacific Islander	33	9.36	7.9	44.59	37.62	42%	39%	18%
Filipino	229	10.07	7.43	47.93	35.38	36%	47%	17%
Hispanic or Latino	3,923	11.38	7.27	54.19	34.6	28%	53%	20%
Black or African American	704	11.4	7.32	54.28	34.84	28%	51%	21%
White	1,405	11.92	7.47	56.77	35.56	26%	49%	25%
Two or more races	115	12.01	7.63	57.18	36.33	29%	45%	26%
English only	3,871	11.68	7.3	55.63	34.77	27%	51%	22%
Initially fluent English proficient	103	9.73	7.32	46.32	34.85	34%	52%	14%
English learner	2,237	10.61	7.49	50.52	35.67	32%	49%	18%
Reclassified fluent English proficient	738	12.54	6.69	59.7	31.85	20%	59%	21%
To be determined	1	NA	NA	NA	NA	NA	NA	NA
English proficiency unknown	6	NA	NA	NA	NA	NA	NA	NA
Not economically disadvantaged	2,390	10.67	7.51	50.8	35.74	32%	49%	19%
Economically disadvantaged	4,566	11.78	7.21	56.08	34.31	26%	53%	21%

#### Table 5.A.3 Demographic Summary for High School

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
Migrant	11	15.55	3.59	74.03	17.09	0%	73%	27%
Nonmigrant	6,945	11.39	7.33	54.23	34.91	28%	52%	21%
Intellectual disability	3,042	11.95	6.81	56.9	32.41	23%	58%	19%
Hearing impairment	66	12.7	6.89	60.46	32.8	20%	56%	24%
Speech or language impairment	43	17.72	3.57	84.39	17.02	2%	40%	58%
Visual impairment	55	7.44	7.73	35.41	36.81	51%	38%	11%
Emotional disturbance	39	17.13	5.03	81.56	23.96	8%	44%	49%
Orthopedic impairment	551	6.85	7.81	32.62	37.19	57%	32%	11%
Other health impairment	218	14.89	6.54	70.88	31.16	13%	45%	42%
Specific learning disability	287	18.12	3.73	86.3	17.76	4%	37%	60%
Deaf-blindness	3	NA	NA	NA	NA	NA	NA	NA
Multiple disabilities	400	6.33	7.16	30.15	34.12	58%	35%	7%
Autism	2,199	11.28	7.15	53.72	34.04	27%	53%	19%
Traumatic brain injury	53	12.42	7.33	59.12	34.91	25%	53%	23%
Not classified	0	NA	NA	NA	NA	NA	NA	NA
American Indian or Alaska Native (Primary ethnicity—Not economically disadvantaged)	20	10.6	8.74	50.48	41.61	40%	30%	30%
Asian (Primary ethnicity—Not economically disadvantaged)	244	11.07	7.13	52.71	33.96	27%	55%	18%
Native Hawaiian or Other Pacific Islander (Primary ethnicity—Not economically disadvantaged)	12	8.17	6.55	38.89	31.18	42%	50%	8%

Student Group	Number Tested	Mean Raw Score	SD of Raw Scores	Mean Percent Correct	SD of Percent Correct	Percent in Preliminary Category 1	Percent in Preliminary Category 2	Percent in Preliminary Category 3
Filipino (Primary ethnicity—Not economically disadvantaged)	147	9.89	7.53	47.1	35.85	38%	44%	18%
Hispanic or Latino (Primary ethnicity—Not economically disadvantaged)	840	10.23	7.44	48.72	35.42	33%	50%	17%
Black or African American (Primary ethnicity—Not economically disadvantaged)	212	10.09	7.05	48.05	33.58	33%	55%	12%
White (Primary ethnicity—Not economically disadvantaged)	837	11.33	7.72	53.96	36.76	30%	46%	23%
Two or more races (Primary ethnicity—Not economically disadvantaged)	60	11.07	7.6	52.7	36.2	32%	48%	20%
American Indian or Alaska Native (Primary ethnicity— Economically disadvantaged)	38	12.95	6.99	61.65	33.27	21%	50%	29%
Asian (Primary ethnicity—Economically disadvantaged)	207	10.24	7.04	48.75	33.52	30%	58%	12%
Native Hawaiian or Other Pacific Islander (Primary ethnicity—Economically disadvantaged)	21	10.05	8.66	47.85	41.22	43%	33%	24%
Filipino (Primary ethnicity—Economically disadvantaged)	82	10.38	7.28	49.42	34.68	32%	52%	16%
Hispanic or Latino (Primary ethnicity—Economically disadvantaged)	3,083	11.69	7.19	55.68	34.23	26%	53%	20%
Black or African American (Primary ethnicity— Economically disadvantaged)	492	11.96	7.36	56.97	35.07	26%	50%	24%
White (Primary ethnicity—Economically disadvantaged)	568	12.79	7	60.9	33.33	20%	52%	27%
Two or more races (Primary ethnicity—Economically disadvantaged)	55	13.04	7.6	62.08	36.17	25%	42%	33%

CAASPP CAA for Science Technical Report | 2017–18 Administration

# Chapter 6: Analyses

This chapter summarizes the item- and test-level statistics from the analyses conducted for the California Alternate Assessment (CAA) for Science administered during the 2017–18 California Assessment of Student Performance and Progress administration.

## 6.1. Sample Used for the Analyses

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

For the 2017–18 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 both classical item analysis and differential item functioning (DIF) analyses.

## 6.2. Classical Item Analyses

Classical item analyses were used to evaluate the second-year pilot 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 correlation indices. Flagging rules based on these statistics identify 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.

## 6.2.1. 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 dichotomous item is:

$$p-value_{dich} = \frac{\sum X_{ic}}{N_i}$$
(6.1)

where,

 $X_{ic}$  is the score received for a given dichotomous item *i* for student *j*, and

 $N_i$  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.

(6.3)

For polytomous items, the *p*-value is defined as:

$$p-value_{poly} = \frac{\sum X_{ij}}{N_i \times Max(X_i)}, \qquad (6.2)$$

where,

 $X_{ij}$  is the score received for a given polytomous item *i* for student *j*,

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

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

#### 6.2.2. 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 PTs. The estimation formula is

$$r_{polyreg} = \frac{\widehat{\beta}s_{tot}}{\sqrt{\widehat{\beta}^2 s_{tot}^2 + 1}}$$

where,

 $\hat{\beta}$  is the estimated slope of the linear regression of the unobservable continuous variable (assumed to account for the item response) on the criterion, and

 $S_{tot}$  is the standard deviation (SD) of the criterion (the students' total raw score).

There is a separate regression for each item score except the lowest, but all regressions for an item are assumed to have a common slope,  $\beta$ . For a polytomous item with *k* possible score values, there are *k*-1 regressions. Beta ( $\beta$ ) 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 itemtotal correlations below 0.20 are flagged for review.

#### 6.2.3. 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 three percent) of students obtaining any possible item score are flagged for further review. Such items may have problems with the item stem or the scoring rubric.

#### 6.2.4. 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 (r-polyserial) is below 0.20.
- 4. The omit rate is above 5 percent for dichotomous items or above 15 percent for polytomous items.
- 5. The percentage is low, with less than three percent of the students at any possible score level.

Refer to Note 1 of <u>appendix 6.A</u> for the flagging symbols, descriptions, and their criteria.

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

#### 6.2.5. Classical Item Analysis Results Summary

This subsection presents tables of the classical item analysis results for the 2017–18 second-year pilot items. Table 6.1 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.

Detailed results of the item analyses for each item by grade are presented in <u>appendix 6.A</u>. The classical item analyses in <u>appendix 6.A</u> are based on the contents of the data file available in June 2018. The item statistics, including AIS, *p*-value, polyserial correlation, statistical flagging criteria, and item type are listed in those tables. The distribution of item scores on each polytomous item is presented in <u>Table 6.A.2</u> through <u>Table 6.A.4</u>.

Test	Number of Students	Number of Items	Number of Points	Mean <i>p</i> -value	Minimum <i>p</i> -value	Maximum <i>p</i> -value	Mean Item-Total Polyserial Correlation	
Grade 5	3,901	19	21	0.58	0.39	0.73	0.59	
Grade 8	3,865	13	27	0.62	0.43	0.76	0.68	
High School	5,559	16	21	0.67	0.50	0.77	0.70	

#### Table 6.1 Classical Item Statistics

## 6.3. Test Completion

Completion rates indicate the proportion of students who complete a specified number of items 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 <u>appendix 6.B</u>.

Data used in <u>appendix 6.B</u> is based on all tested students in the full student population.

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

The frequency distribution of scores on each test and the mean raw score for the tests overall and for each embedded PT are used to evaluate the performance of the second-year pilot tests and PTs. The mean raw scores are provided in <u>Table 6.2</u> and are based on the item analysis sample. The distributions of raw scores are provided in <u>appendix 6.C</u>. Data used in <u>appendix 6.C</u> includes all tested students in the full student population. The "NA" notation in the tables in <u>appendix 6.C</u> indicates that the number of points were not possible for the PT.

Module	Number of Students	Number of Items	Maximum Number of Points	Mean Raw Score	SD Raw Score	Minimum Raw Score	Maximum Raw Score
Grade 5 Total:	3,901	19	21	12.4	4.8	0	21
Grade 5 PT 1 (Sun and Shadows)	3,901	7	9	5.0	2.3	0	9
Grade 5 PT 2 (Physical Changes)	3,901	6	6	3.6	1.7	0	6
Grade 5 PT 3 (Weather Conditions)	3,901	6	6	3.7	1.7	0	6

#### Table 6.2 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
Grade 8 Total:	3,865	13	27	17.5	6.6	0	27
Grade 8 PT 1 (Water Cycle)	3,865	6	8	4.8	2.2	0	8
Grade 8 PT 2 (Bioenergy)	3,865	4	8	4.8	2.4	0	8
Grade 8 PT 3 (Cells)	3,865	3	11	7.9	3.0	0	11
High School Total:	5,559	16	21	14.1	5.5	0	21
High School PT 1 (Molecules)	5,559	3	8	5.5	2.2	0	8
High School PT 2 (Force and Motion)	5,559	7	7	5.1	2.1	0	7
High School PT 3 (Erosion)	5,559	6	6	3.5	1.9	0	6

## 6.5. Differential Item Functioning (DIF)

DIF analyses were conducted for 2017–18 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 else other than the intended construct (i.e., possible evidence of bias). It is important, however, to recognize that item performance differences flagged for DIF might be related to actual differences in relevant knowledge or skills between student groups (i.e., impact) or statistical Type I error, which might falsely find DIF in an item. As a result, DIF analysis is used mainly as a statistical tool to identify *potential* item bias. Subsequent reviews by content experts and bias and sensitivity experts are required to determine the source and meaning of performance differences.

## 6.5.1. 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., White students). The common odds ratio is estimated across the total raw score using the formula in Equation 6.4 (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.

$$\alpha_{MH} = \frac{\left(\sum_{m} R_{rm} \frac{W_{fm}}{N_{tm}}\right)}{\left(\sum_{m} R_{fm} \frac{W_{rm}}{N_{tm}}\right)}$$
(6.4)

where,

m indexes the score categories,

 $R_{rm}$  is the number of students in the reference group who answer the item correctly,

 $W_{fm}$  is the number of students in the focal group who answer the item incorrectly,

 $N_{tm}$  is the total number of students,

 $R_{fm}$  is the number of students in the focal group who answer the item correctly, and

 $W_{rm}$  is the number of students in the reference 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 following formula (Holland & Thayer, 1988):

M H D-D I F=-2.351n  $\left[\alpha_{MH}\right]$ 

(6.5)

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 item are differentially easier for the reference group.

## 6.5.2. 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 following formula:

$$SMD = \frac{\sum_{m=1}^{M} N_{fm} \times (E_f(Y \mid X = m) - E_r(Y \mid X = m)))}{\sum_{m=1}^{M} N_{fm}}$$
(6.6)

where,

X is the criterion score,

Y is the item score,

M is the number of score categories on X,

 $N_{rm}$  is the number of students in the reference group in score category m,

 $N_{fm}$  is the number of students in the focal group in score category m,

 $E_r$  is the expected item score for the reference group, and

 $E_t$  is the expected item score for the focal group.

A positive SMD value means that, conditional on the criterion score, the focal group has a higher mean item score than the reference group. In contrast, a negative SMD value means that, conditional upon the criterion score, the focal group has a lower mean item score than the reference group.

## 6.5.3. DIF Categories and Definitions

Based on the DIF statistics and significance tests, items are classified into three categories and assigned values of A, B, or 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 flagging criteria for dichotomous items are presented in <u>Table 6.3</u>; the flagging criteria for polytomous items are provided in <u>Table 6.4</u>.

DIF	
Category	Criteria
A (negligible)	<ul> <li>Absolute value of MH D-DIF is not significantly different from zero, or is less than one.</li> </ul>
	<ul> <li>Positive values are classified as "A+" and negative values as "A"</li> </ul>
B (moderate)	<ul> <li>Absolute value of MH D-DIF is significantly different from zero but not from one, and is at least one; OR</li> <li>Absolute value of MH D-DIF is significantly different from one, but is less than 1.5.</li> </ul>
	<ul> <li>Positive values are classified as "B+" and negative values as "B"</li> </ul>
C (large)	<ul> <li>Absolute value of MH D-DIF is significantly different from one, and is at least 1.5.</li> </ul>
	<ul> <li>Positive values are classified as "C+" and negative values as "C"</li> </ul>

#### Table 6.3 DIF Categories for Dichotomous Items

Table 6.4 DIF Categories for Polytomous Iten	ns
--	----

DIF Category	Criteria
A (negligible)	<ul> <li>Mantel Chi-square p value &gt; 0.05 or  SMD/SD  ≤ 0.17</li> </ul>
B (moderate)	<ul> <li>Mantel Chi-square <i>p</i> value &lt; 0.05 or 0.17&lt;  SMD/SD  ≤ 0.25</li> </ul>
C (large)	<ul> <li>Mantel Chi-square p value &lt; 0.05 or  SMD/SD  &gt; 0.25</li> </ul>

**Note:** *SMD* = standardized mean difference; SD = total group standard deviation of item score

DIF analyses were conducted on each CAA for Science for designated comparison groups. Groups were defined on the basis of demographic variables, including gender, race or ethnicity, and primary disabilities, if the number of students in the group was sufficient. These comparison groups are specified in <u>Table 6.5</u>.

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

Table 6.5 S	Student Groups	for DIF Comparison
-------------	----------------	--------------------

\* DIF analysis was not performed due to insufficient sample sizes for all three CAA for Science Tests.

Note the following about specific DIF analyses:

- The Filipino versus White DIF analysis was not performed for grade five due to insufficient sample size.
- The Speech or Language Impairment versus Intellectual Disability DIF analysis was not performed due to insufficient sample sizes for grade eight and high school.
- The grade ten versus grade eleven DIF analysis was not performed due to insufficient sample sizes for the high school test.

## 6.5.4. Items Exhibiting Significant DIF

The DIF results tables in <u>appendix 6.D</u>, which are based the data file available in June 2018, include the number of items with sufficient sample sizes to be included in the DIF analyses (the *N* column). In addition, "-" indicates that the DIF analysis did not classify any items in the particular DIF category, while "NA" indicates that the DIF analysis was not performed due to insufficient sample size. <u>Table 6.D.8</u> lists the item flagged during the DIF analyses.

## References

- 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.
- Holland, P. W., & Thayer, D. T. (1985). *An alternative definition of the ETS delta scale of item difficulty* (Research Report 85–43). Princeton, NJ: Educational Testing Service.
- Holland, P., & 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.
- 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.
- 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), 321–344.

## **Appendix 6.A: Classical Item Analyses**

Module	Number of Students	Number of Unique Items	Maximum Number of Points	Mean <i>p</i> -value	Lowest Observed <i>p</i> - value	Highest Observed <i>p</i> - value	Mean Item-total Polyserial Correlation
Grade Five Total:	3,901	19	21	0.58	0.39	0.73	0.59
Grade Five PT 1	3,901	7	9	0.53	0.39	0.73	0.56
Grade Five PT 2	3,901	6	6	0.60	0.46	0.68	0.59
Grade Five PT 3	3,901	6	6	0.62	0.48	0.73	0.62
Grade Eight Total:	3,865	13	27	0.62	0.43	0.76	0.68
Grade Eight PT 1	3,865	6	8	0.59	0.43	0.73	0.61
Grade Eight PT 2	3,865	4	8	0.61	0.58	0.66	0.72
Grade Eight PT 3	3,865	3	11	0.69	0.59	0.76	0.75
High School Total:	5,559	16	21	0.67	0.50	0.77	0.70
High School PT 1	5,559	3	8	0.70	0.66	0.75	0.73
High School PT 2	5,559	7	7	0.73	0.62	0.77	0.71
High School PT 3	5,559	6	6	0.58	0.50	0.64	0.67

#### Table 6.A.1 Classical Item Statistics for Each Embedded PT

In <u>Table 6.A.2</u> through <u>Table 6.A.4</u>, 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 <u>Table 6.A.2</u> through <u>Table 6.A.4</u>.

Flag	Description	Criteria
A	Indicates low average item score (AIS) /low <i>p</i> -value (difficult item)	Dichotomous item: <i>p</i> -value < 0.33 Polytomous item: AIS < 30 percent of maximum possible score points
Н	Indicates high AIS/high <i>p</i> -value (easy item)	Dichotomous item: <i>p</i> -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	Polyserial < 0.20
0	Indicates high percent of omits/not responding	Dichotomous item: %omit > 5% Polytomous item: %omit > 15%
L	Indicates few students at one or more score levels (for polytomous items only)	Polytomous item: less than 3% at one or more score levels

						Maximum
	ltem				Omit	Score
Item ID	Use	Flag	AIS	Polyserial	Rate	Points
223-4111	Pilot	[no flag]	0.57	0.60	3%	1
223-4113	Pilot	[no flag]	0.64	0.67	4%	1
223-4115	Pilot	0	0.67	0.56	5%	1
223-4117	Pilot	[no flag]	0.68	0.53	4%	1
223-4119	Pilot	0	0.56	0.57	7%	1
223-4121	Pilot	[no flag]	0.46	0.60	4%	1
223-4131	Pilot	[no flag]	0.73	0.63	2%	1
223-4133	Pilot	[no flag]	0.59	0.57	5%	1
223-4135	Pilot	[no flag]	0.39	0.57	5%	1
223-4137	Pilot	[no flag]	0.47	0.49	3%	1
223-4139	Pilot	[no flag]	0.42	0.45	3%	1
223-4141	Pilot	[no flag]	0.45	0.54	3%	1
223-4143	Pilot	[no flag]	2.01	0.70	2%	3
223-4151	Pilot	[no flag]	0.73	0.69	1%	1
223-4153	Pilot	[no flag]	0.69	0.60	2%	1
223-4155	Pilot	0	0.52	0.60	6%	1
223-4157	Pilot	[no flag]	0.69	0.62	2%	1
223-4159	Pilot	[no flag]	0.48	0.55	3%	1
223-4161	Pilot	[no flag]	0.62	0.63	3%	1

#### Table 6.A.2 Average Item Score and Polyserial Correlation for Each Item: Grade Five

Table 6.A.3	Average Item Sco	e and Polyseria	I Correlation fo	or Each Item:	Grade Eight

	ltem				Omit	Maximum Score
Item ID	Use	Flag	AIS	Polyserial	Rate	Points
223-4185	Pilot	[no flag]	1.86	0.76	1%	3
223-4187	Pilot	[no flag]	0.66	0.66	2%	1
223-4189	Pilot	[no flag]	0.59	0.67	2%	1
223-4191	Pilot	[no flag]	0.50	0.53	2%	1
223-4193	Pilot	[no flag]	0.73	0.67	3%	1
223-4195	Pilot	[no flag]	0.58	0.56	4%	1
223-4197	Pilot	[no flag]	0.43	0.53	5%	1
223-4199	Pilot	[no flag]	0.69	0.67	4%	1
223-4201	Pilot	[no flag]	1.17	0.70	2%	2
223-4203	Pilot	[no flag]	1.15	0.76	2%	2
223-4223	Pilot	L	5.82	0.86	0%	8
223-4225	Pilot	[no flag]	1.51	0.73	2%	2
223-4229	Pilot	[no flag]	1.82	0.68	1%	3

						Maximum
	ltem				Omit	Score
Item ID	Use	Flag	AIS	Polyserial	Rate	Points
223-4259	Pilot	[no flag]	0.76	0.70	2%	1
223-4261	Pilot	[no flag]	0.75	0.70	3%	1
223-4263	Pilot	[no flag]	0.75	0.71	3%	1
223-4265	Pilot	[no flag]	0.77	0.73	2%	1
223-4267	Pilot	[no flag]	0.71	0.72	3%	1
223-4269	Pilot	[no flag]	0.73	0.74	3%	1
223-4271	Pilot	[no flag]	0.62	0.68	4%	1
223-4273	Pilot	[no flag]	0.75	0.69	2%	1
223-4275	Pilot	[no flag]	0.66	0.69	3%	1
223-4277	Pilot	[no flag]	0.60	0.64	3%	1
223-4279	Pilot	[no flag]	0.50	0.63	4%	1
223-4281	Pilot	[no flag]	0.57	0.67	4%	1
223-4283	Pilot	[no flag]	0.57	0.69	4%	1
223-4285	Pilot	[no flag]	0.60	0.70	4%	1
223-4287	Pilot	[no flag]	0.64	0.69	4%	1
223-4313	Pilot	[no flag]	4.12	0.81	1%	6

### Table 6.A.4 Average Item Score and Polyserial Correlation for Each Item: High School

Item ID	Item Use	Max Points	Score 0	Score 1	Score 2	Score 3	Blank
223-4143	Pilot	3	10%	28%	6%	53%	2%

							•			•		
	ltem	Мах	Score									
Item ID	Use	Points	0	1	2	3	4	5	6	7	8	Blank
223-4185	Pilot	3	14%	22%	25%	38%	NA	NA	NA	NA	NA	1%
223-4201	Pilot	2	13%	52%	32%	NA	NA	NA	NA	NA	NA	2%
223-4203	Pilot	2	26%	28%	43%	NA	NA	NA	NA	NA	NA	2%
223-4223	Pilot	8	2%	3%	5%	7%	13%	8%	11%	18%	33%	0%
223-4225	Pilot	2	9%	27%	62%	NA	NA	NA	NA	NA	NA	2%
223-4229	Pilot	3	15%	31%	9%	44%	NA	NA	NA	NA	NA	1%

Table 6.A.6 Distribution of Item Scores for the Polytomous Items: Grade Eight

 Table 6.A.7 Distribution of Item Scores for the Polytomous Item: Grade High School

Item ID	ltem Use	Max Points	Score 0	Score 1	Score 2	Score 3	Score 4	Score 5	Score 6	Blank
223-4313	Pilot	6	4%	4%	10%	8%	34%	6%	33%	1%

## **Appendix 6.B: Completion Rates**

**Note:** In <u>Table 6.B.1</u>, the number of students completing the performance task (PT) refers to the number of students who attempted the PT.

Grade Level	Embedded PT	Number of Students Registered	Number of Students Completing the PT	Percent Completing the PT
Five	1	5,766	4,800	83%
Five	2	5,766	4,744	82%
Five	3	5,766	4,776	83%
Eight	1	5,807	4,585	79%
Eight	2	5,807	4,551	78%
Eight	3	5,807	4,545	78%
High School	1	10,234	7,101	69%
High School	2	10,234	7,007	68%
High School	3	10,234	7,017	69%

#### Table 6.B.1 Completion Rate

 Table 6.B.2 Percentage of Students Completing the PTs

Grade Level	No PTs Completed	1 PT Completed	2PTs Completed	3 PTs Completed	Number of Students Registered
Five	16%	2%	1%	82%	5,766
Eight	20%	1%	0%	78%	5,807
High School	30%	1%	1%	68%	10,234
Ten	55%	3%	1%	41%	186
Eleven	34%	1%	1%	65%	1,789
Twevel	29%	1%	1%	69%	8,259

**Note:** Number of Students is based on the total number of students registered in each grade.

## Appendix 6.C: Distribution of Raw Scores Overall and by Embedded Performance Task (PT)

#### Table 6.C.1 Distribution of Students for Total Score and PT Scores for Grade Five

Raw Score	Number of Students (Total)	Percentage of Students (Total)	Number of Students (PT 1, Sun and Shadows)	Percentage of Students (PT 1, Sun and Shadows)	Number of Students (PT 2, Physical Changes)	Percentage of Students (PT 2, Physical Changes)	Number of Students (PT 3, Weather Conditions)	Percentage of Students (PT 3, Weather Conditions)
0	646	14%	757	16%	945	20%	839	18%
1	62	1%	217	5%	302	6%	397	8%
2	67	1%	337	7%	561	12%	522	11%
3	81	2%	462	10%	771	16%	665	14%
4	106	2%	561	12%	821	17%	781	17%
5	119	3%	592	13%	678	14%	740	16%
6	143	3%	635	13%	634	13%	768	16%
7	193	4%	528	11%	NA	NA	NA	NA
8	208	4%	376	8%	NA	NA	NA	NA
9	238	5%	247	5%	NA	NA	NA	NA
10	274	6%	NA	NA	NA	NA	NA	NA
11	296	6%	NA	NA	NA	NA	NA	NA
12	300	6%	NA	NA	NA	NA	NA	NA
13	267	6%	NA	NA	NA	NA	NA	NA
14	300	6%	NA	NA	NA	NA	NA	NA
15	288	6%	NA	NA	NA	NA	NA	NA
16	244	5%	NA	NA	NA	NA	NA	NA
17	253	5%	NA	NA	NA	NA	NA	NA
18	212	4%	NA	NA	NA	NA	NA	NA
19	183	4%	NA	NA	NA	NA	NA	NA
20	129	3%	NA	NA	NA	NA	NA	NA
21	103	2%	NA	NA	NA	NA	NA	NA

le 6.C.2	Distrib	ution o	f Stud	ents fo	or lota	I Score	e and H		res to	r G
	Raw Score	Number of Students (Total)	Percentage of Students (Total)	Number of Students (PT 1, Water Cycle)	Percentage of Students (PT 1, Water Cycle)	Number of Students (PT 2, Bioenergy)	Percentage of Students (PT 2, Bioenergy)	Number of Students (PT 3, Cells)	Percentage of Students (PT 3, Cells)	
	0	590	13%	749	17%	750	17%	661	15%	
	1	38	1%	225	5%	241	5%	69	2%	
	2	30	1%	331	7%	375	8%	100	2%	
	3	38	1%	478	11%	460	10%	140	3%	
	4	42	1%	547	12%	493	11%	199	4%	
	5	42	1%	569	13%	479	11%	296	7%	
	6	57	1%	570	13%	519	12%	368	8%	
	7	74	2%	571	13%	535	12%	341	8%	
	8	102	2%	472	10%	660	15%	344	8%	
	9	109	2%	NA	NA	NA	NA	389	9%	
	10	155	3%	NA	NA	NA	NA	566	13%	
	11	134	3%	NA	NA	NA	NA	1,039	23%	
	12	158	4%	NA	NA	NA	NA	NA	NA	
	13	169	4%	NA	NA	NA	NA	NA	NA	
	14	148	3%	NA	NA	NA	NA	NA	NA	
	15	167	4%	NA	NA	NA	NA	NA	NA	
	16	209	5%	NA	NA	NA	NA	NA	NA	
	17	189	4%	NA	NA	NA	NA	NA	NA	
	18	181	4%	NA	NA	NA	NA	NA	NA	
	19	166	4%	NA	NA	NA	NA	NA	NA	
	20	190	4%	NA	NA	NA	NA	NA	NA	
	21	195	4%	NA	NA	NA	NA	NA	NA	
	22	238	5%	NA	NA	NA	NA	NA	NA	
	23	193	4%	NA	NA	NA	NA	NA	NA	
	24	246	5%	NA	NA	NA	NA	NA	NA	
	25	215	5%	NA	NA	NA	NA	NA	NA	
	26	232	5%	NA	NA	NA	NA	NA	NA	
	27	205	5%	NA	NA	NA	NA	NA	NA	

## Table 6.C.2 Distribution of Students for Total Score and PT Scores for Grade Eight

Raw Score	Number of Students (Total)	Percentage of Students (Total)	Number of Students (PT 1, Molecules)	Percentage of Students (PT 1, Molecules)	Number of Students (PT 2, Force and Motion)	Percentage of Students (PT 2, Force and Motion)	Number of Students (PT 3, Erosion)	Percentage of Students (PT 3, Erosion)
0	1,314	19%	1,464	21%	1,550	22%	1,901	27%
1	90	1%	203	3%	233	3%	521	7%
2	105	2%	287	4%	383	6%	754	11%
3	74	1%	360	5%	500	7%	844	12%
4	97	1%	582	8%	550	8%	948	14%
5	117	2%	864	12%	668	10%	853	12%
6	140	2%	1,133	16%	942	14%	1,135	16%
7	178	3%	568	8%	2,130	31%	NA	NA
8	215	3%	1,495	21%	NA	NA	NA	NA
9	221	3%	NA	NA	NA	NA	NA	NA
10	236	3%	NA	NA	NA	NA	NA	NA
11	239	3%	NA	NA	NA	NA	NA	NA
12	260	4%	NA	NA	NA	NA	NA	NA
13	332	5%	NA	NA	NA	NA	NA	NA
14	318	5%	NA	NA	NA	NA	NA	NA
15	333	5%	NA	NA	NA	NA	NA	NA
16	413	6%	NA	NA	NA	NA	NA	NA
17	399	6%	NA	NA	NA	NA	NA	NA
18	445	6%	NA	NA	NA	NA	NA	NA
19	450	6%	NA	NA	NA	NA	NA	NA
20	425	6%	NA	NA	NA	NA	NA	NA
21	555	8%	NA	NA	NA	NA	NA	NA

## Appendix 6.D: Differential Item Functioning (DIF) Analysis: Number and Percentage of Items in Each DIF Category

		Tab	ole 6.	D.1 DIF	for (	Grade F	ive			
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+	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
B+	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
A+	12	63%	10	53%	12	63%	NA	NA	8	42%
A-	7	37%	9	47%	7	37%	NA	NA	11	58%
B-	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
C-	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	19	100%	NA	NA
Operational Items Total	19 <i>^</i>	100%	19	100%	19	100%	19	100%	19	100%

Table 6.D.1 DIF for Grade Five

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+	NA	NA	NA	NA	NA	NA	NA	NA	1	5%	NA	NA
B+	NA	NA	1	5%	NA	NA	NA	NA	2	11%	2	11%
A+	7	37%	9	47%	11	58%	9	47%	7	37%	6	32%
A-	12	63%	8	42%	7	37%	9	47%	9	47%	11	58%
В-	NA	NA	1	5%	1	5%	1	5%	NA	NA	NA	NA
C-	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Operational Items Total	19	100%	19	100%	19	100%	19	100%	19	100%	19	100%

### Table 6.D.2 DIF for Grade Five (Continued)

## Table 6.D.3 DIF for 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+	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
B+	NA	NA	1	8%	NA	NA	1	8%	NA	NA
A+	6	46%	6	46%	5	38%	4	31%	7	54%
A	7	54%	6	46%	8	62%	8	62%	6	46%
В-	N/	A NA	NA	NA	NA	NA	NA	NA	NA	NA
C-	N/	A NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	N	A NA	NA	NA	NA	NA	NA	NA	NA	NA
Operational Items Total	13	100%	13	100%	13 1	00%	<b>13</b> 1	00%	13	100%

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+	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
B+	NA	NA	NA	NA	NA	NA	NA	NA	2	15%	NA	NA
A+	6	46%	6	46%	5	38%	9	69%	6	46%	NA	NA
A-	7	54%	7	54%	7	54%	4	31%	5	38%	NA	NA
B-	NA	NA	NA	NA	1	8%	NA	NA	NA	NA	NA	NA
C-	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	13	100%
Operational Items Total	13	100%	13	100%	13	100%	13	100%	13	100%	13	100%

### Table 6.D.4 DIF for Grade Eight (Continued)

#### Table 6.D.5 DIF for High School

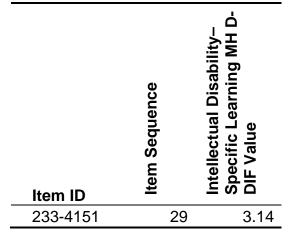
DIF Category	Male-Female N	Male-Female Pct	White-African Amer N	White-African Amer Pct	White–Asian N	White–Asian Pct	White–Filipino N	White-Filipino Pct	White-Hispanic N	White-Hispanic Pct
C+	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
B+	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
A+	5	31%	9	56%	5	31%	7	44%	10	63%
A-	11	69%	7	44%	11	69%	9	56%	6	38%
B-	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
C-	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Operational Items Total	16 <sup>-</sup>	100%	16	100%	16	100%	16	100%	16	100%

DIF Category	Intellectual Disability– Autism N Intellectual Disability–		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+	NA NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
B+	NA NA	A NA	NA	NA	NA	NA	NA	1	6%	NA	NA
A+	7 449	6 8	50%	9	56%	10	63%	8	50%	NA	NA
A-	9 569	6 8	50%	7	44%	6	38%	7	44%	NA	NA
В-	NA NA	A NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
C-	NA N	A NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
NA	NA N	A NA	NA	NA	NA	NA	NA	NA	NA	16	100%
Operational Items Total	16 1009	% 16	100%	16	100%	16	100%	16	100%	16	100%

## Table 6.D.6 DIF for High School (Continued)

Table 6.D.7	<b>DIF for High School</b>	(Continued)
-------------	----------------------------	-------------

DIF Category	Grade 10 N	Grade 10 Pct	Grade 12 N	Grade 12 Pct
C+	NA	NA	NA	NA
B+	NA	NA	NA	NA
A+	NA	NA	9	56%
A-	NA	NA	7	44%
B-	NA	NA	NA	NA
C-	NA	NA	NA	NA
NA	16	100%	NA	NA
Operational Items Total	16	100%	16	100%



#### Table 6.D.8 Item Exhibiting Significant DIF—Grade Five

## Chapter 7: Surveys

Three separate surveys were developed and administered to collect additional information on the second year of California Alternate Assessment (CAA) for Science pilot testing in 2017–18: a student engagement survey completed by test examiners about each student's level of engagement with each embedded performance task (PT); a student survey answered by students about their reactions to each embedded PT; and a test examiner survey completed by test examiners to solicit their feedback on the CAA for Science administration overall.

## 7.1. Survey Design and Questionnaire Development

These 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.

## 7.1.1. Student Survey and Student Engagement Survey

Student survey responses, which are provided by the student, and student engagement survey responses, provided by test examiners, were collected from local educational agencies (LEAs) for every embedded PT administered to every student. After a PT was administered to the student, his or her test examiner would then administer a student survey consisting of two questions. Thereafter, the test examiner would answer two questions regarding the student's engagement with the PT he or she had just administered.

The purposes of the student survey and student engagement survey were to collect

- basic information about students' experience with the embedded PT just administered, and
- information about the mode of communication and level of engagement with the embedded PT just administered.

The first of two questions in the student survey asked the student how he or she felt about taking the PT. The second asked if the student had enough time to complete the PT. The student engagement survey asked the test examiner about the mode of communication used by the student, as well as the student's level of engagement with the PT.

The student survey and the student engagement survey were included as the last section of each downloadable embedded PT document. After marking responses to the surveys on an Answer Recording Document, test examiners later recorded the results of these two surveys for each student with the results of each embedded PT in the Data Entry Interface.

## 7.1.2. 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 second-year pilot administration and assessment processes overall. This survey was linked on the California Assessment of Student Performance and Progress Portal at <a href="http://www.caaspp.org/">http://www.caaspp.org/</a> and hosted on SurveyGizmo.com, a website with survey-creation and hosting services.

## 7.2. Student Survey Administration

For the student survey, test examiners were asked to direct the student's attention to a graphic, read the survey question aloud, and have students indicate their answer by pointing or communicating in the mode appropriate for them. The test examiner would then record the student's response on the Answer Recording Document.

<u>Figure 7.1</u> shows the first question in the student survey. Students were asked, "How do you feel about the test your teacher gave you?" The three possible responses were:

- а. Нарру
- b. Sad
- c. Confused

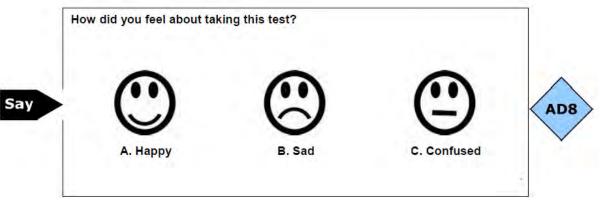
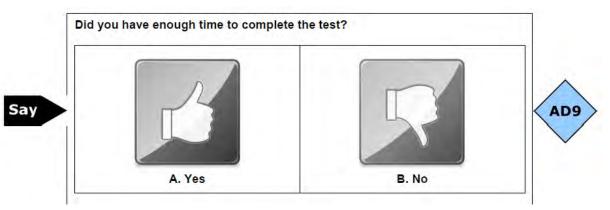
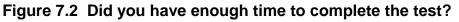


Figure 7.1 How did you feel about taking this test?

Figure 7.2 shows the second question in the student survey. Students were asked, "Did you have enough time to complete the test?" The two possible responses were:

- a. Yes
- b. No





The test examiner would also answer the student engagement survey after testing was completed. The questions pertained to the mode of communication the student used and the level of engagement the student displayed as the task was administered.

## 7.3. Student Survey Results

Results for the student survey and student engagement survey are shown in the next subsection, 7.3.1. Results from the optional test examiner survey are shown in subsection <u>7.4 Test Examiner Survey Results</u>.

## 7.3.1. Summary of Survey Results Regarding Student Experiences

Generally, most students indicated "Happy" when queried "How did you feel about taking this test?" Across all PTs, a range from 12.1 percent to 14.7 percent indicated no response to this question. A range of 3.0 percent to 4.5 percent indicated "Sad" in response to this question. A range of 11.1 percent to 14.4 percent indicated "Confused." Responses are shown in <u>Table 7.1</u>.

#### Table 7.1 Student Survey Responses—How did you feel about taking this test?

Frakedded Derfermen es Tesk	Number of Respondents	Henry	Cod	Confused	No Response
Embedded Performance Task			Sad	Confused	
Grade Five, PT 1: Sun and Shadows	4,569	68.0%	4.2%	14.4%	13.3%
Grade Five, PT 2: Physical Changes	4,514	71.3%	4.2%	11.1%	13.4%
Grade Five, PT 3: Weather Conditions	4,528	71.2%	4.5%	11.6%	12.7%
Grade Eight, PT 1: Water Cycle	4,390	70.2%	3.8%	13.6%	12.3%
Grade Eight, PT 2: Bioenergy	4,382	72.0%	3.2%	11.8%	12.9%
Grade Eight, PT 3: Cells	4,379	73.7%	3.1%	11.1%	12.1%
High School, PT 1: Molecules	6,555	68.5%	3.2%	13.6%	14.7%
High School, PT 2: Force and Motion	6,534	70.3%	3.0%	12.5%	14.2%
High School, PT 3: Erosion	6,502	68.4%	3.4%	14.3%	13.9%

For each PT administered, more than 85 percent of students indicated that they did have enough time to complete the test. Across all embedded PTs, a range from 8.4 percent to 10.7 percent of students indicated no response to this question. Responses are shown in Table 7.2.

Table 7.2 Student Survey Responses—Did you have enough time to complete
the test?

Embedded Performance Task	Number of Respondents	Yes	No	No Response
Grade Five, PT 1: Sun and Shadows	4,560	88.0%	3.0%	9.0%
Grade Five, PT 2: Physical Changes	4,515	87.8%	2.9%	9.3%
Grade Five, PT 3: Weather Conditions	4,521	87.8%	2.7%	9.4%

Embedded Performance Task	Number of Respondents	Yes	Νο	No Response
Grade Eight, PT 1: Water Cycle	4,387	89.5%	2.1%	8.4%
Grade Eight, PT 2: Bioenergy	4,379	88.9%	2.0%	9.1%
Grade Eight, PT 3: Cells	4,377	88.8%	2.2%	9.0%
High School, PT 1: Molecules	6,560	87.0%	2.3%	10.7%
High School, PT 2: Force and Motion	6,518	87.5%	2.3%	10.3%
High School, PT 3: Erosion	6,495	87.4%	2.4%	10.1%

Test examiners were directed to answer two questions regarding the student's level of engagement. The first question asked gave six answer options, including "Other," for the test examiner to indicate mode of engagement. According to the answers to these questions, the majority of students used "Gestures or pointing" or "verbal responses" to answer questions. A range of 4.7 percent to 6.4 percent used "mouse, touchscreen, and/or computer keyboard." A range of 1.5 percent to 3.6 percent used assistive/augmentative communication devices. A range of 1.5 percent to 3.9 percent responses indicated eye gaze was used. Responses are shown in <u>Table 7.3</u>.

## Table 7.3 Student Engagement Survey Responses—Select the mode(s) of communication used by the student on this performance task. (Select all that apply)

Embedded Performance Task	Number of Respondents	Mouse, touchscreen, and/or computer keyboard	Verbal Response	Gestures or pointing	Assistive/augmentative communication device	Eye gaze	Other – please specify
Grade Five, Task 1: Sun and Shadows	4,383	6.1%	77.8%	84.2%	1.6%	3.9%	6.5%
Grade Five, Task 2: Physical Changes	4,335	6.4%	80.0%	80.8%	2.2%	3.8%	6.8%
Grade Five, Task 3: Weather Conditions	4,359	5.9%	78.6%	84.1%	1.5%	3.9%	6.4%
Grade Eight, Task 1: Water Cycle	4,229	4.9%	78.9%	79.0%	3.6%	2.6%	6.5%
Grade Eight, Task 2: Bioenergy	4,224	4.7%	77.9%	81.1%	2.9%	2.2%	6.4%
Grade Eight, Task 3: Cells	4,223	4.8%	78.0%	80.9%	2.9%	2.3%	6.3%

Embedded Performance Task	Number of Respondents	Mouse, touchscreen, and/or computer keyboard	Verbal Response	Gestures or pointing	Assistive/augmentative communication device	Eye gaze	Other – please specify
High School, Task 1: Molecules	6,301	5.5%	76.9%	64.0%	3.0%	1.5%	9.1%
High School, Task 2: Force and Motion	6,279	5.0%	78.0%	60.2%	3.4%	1.8%	9.0%
High School, Task 3: Erosion	6,259	5.0%	77.7%	57.8%	3.2%	1.7%	9.2%

If "Other" was selected for the previous question, test examiners were asked to specify. Table 7.4 through Table 7.6 show categories of the responses for each grade. The 10 response modes indicated are as follows:

- 1. American Sign Language (ASL)
- 2. Picture exchange communication
- 3. Letter choices
- 4. Picture or labels on diagrams
- 5. Yes/No icons
- 6. Response cards
- 7. Options picked from written responses
- 8. Objects used to offer a multiple-choice option
- 9. Spanish interpreters
- 10. Partner-assisted scanning

Most Common Other Mode	Grade 5 PT 1 (Sun and Shadows)— Number of Respondents	Grade 5 PT 1 (Sun and Shadows)— Percent	Grade 5 PT 2 (Physical Changes )— Number of Respondents	Grade 5 PT 2 (Physical Changes)— Percent	Grade 5 PT 3 (Weather Conditions)— Number of Respondents	Grade 5 PT 3 (Weather Conditions)— Percent
ASL	17	50.0%	21	58.3%	19	61.3%
Picture exchange communication	5	14.7%	7	19.4%	6	19.4%
Letter choices	4	11.8%	4	11.1%	4	12.9%
Placed pictures or labels on diagram	3	8.8%	NA	NA	NA	NA
Yes/No icons	2	5.9%	1	2.8%	1	3.2%
Response cards	NA	NA	2	5.6%	NA	NA
Spanish interpreter	1	2.9%	1	2.8%	1	3.2%

## Table 7.4 Student Engagement Survey Responses—"Other" Mode(s) of<br/>Communication: Most Common Responses for Grade Five

Most Common Other Mode	Grade 8 PT 1 (Water Cycle)— Number of Respondents	Grade 8 PT 1 (Water Cycle)— Percent	Grade 8 PT 2 (Bioenergy)— Number of Respondents	Grade 8 PT 2 (Bioenergy)— Percent	Grade 8 PT 3 (Cells)— Number of Respondents	Grade 8 PT 3 (Cells)— Percent
ASL	33	60.0%	32	53.3%	30	66.7%
Picture exchange communication	NA	NA	NA	NA	NA	NA
Letter choices	1	1.8%	NA	NA	NA	NA
Placed pictures or labels on diagram	21	38.2%	28	46.7%	15	33.3%

## Table 7.5 Student Engagement Survey Responses—"Other" Mode(s) of<br/>Communication: Most Common Responses for Grade Eight

## Table 7.6 Student Engagement Survey Responses—"Other" Mode(s) of<br/>Communication: Common Responses for High School

Most Common Other Mode	High School PT 1 (Molecules)—Number of	High School PT 1 (Molecules)—Percent	High School PT 2 (Force and Motion)—Number of	High School PT 2 (Force and Motion)—Percent	High School PT 3 (Erosion)— Number of Respondents	High School PT 3 (Erosion)— Percent
ASL	71	66.4%	50	63.3%	50	68.5%
Picture exchange communication	8	7.5%	7	8.9%	5	6.8%
Letter choices	3	2.8%	6	7.6%	5	6.8%
Yes/No icons	2	1.9%	3	3.8%	3	4.1%
Response cards	12	11.2%	NA	NA	NA	NA
Picked from written responses	5	4.7%	6	7.6%	2	2.7%
Object used to offer multiple-choice option	4	3.7%	5	6.3%	6	8.2%
Spanish interpreter	1	0.9%	1	1.3%	1	1.4%

The last question asked in the student engagement survey pertained to the student's level of engagement with the task. As displayed in <u>Table 7.7</u>, for each task, more than half of the respondents indicated that the student was "Fully Engaged." A range of 17.6 percent to 25.6 percent of responses indicated moderate engagement. A range of 19.0 percent to 23.2 percent of responses indicated minimal engagement.

with this performance task you just administered?						
Embedded Performance Task	Number of Respondents	Fully Engaged	Moderately Engaged	Minimally Engaged		
Grade Five, Task 1: Sun and Shadows	4,502	51.4%	25.6%	23.0%		
Grade Five, Task 2: Physical Changes	4,459	51.8%	25.0%	23.2%		
Grade Five, Task 3: Weather Conditions	4,467	56.1%	23.1%	20.9%		
Grade Eight, Task 1: Water Cycle	4,346	58.7%	21.6%	19.7%		
Grade Eight, Task 2: Bioenergy	4,345	58.5%	21.6%	19.9%		
Grade Eight, Task 3: Cells	4,345	61.2%	19.8%	19.0%		
High School, Task 1: Molecules	6,517	63.1%	17.6%	19.3%		
High School, Task 2: Force and Motion	6,468	62.0%	18.3%	19.7%		
High School, Task 3: Erosion	6,458	59.9%	19.0%	21.1%		

## Table 7.7 Student Engagement Survey Responses—How engaged was the student with this performance task you just administered?

## 7.4. Test Examiner Survey Results

Twenty questions were asked on the test examiner survey. The first question asked whether or not the respondent was a test examiner for the 2017–18 CAA for Science. Of the 883 responses to the first question, 784 indicated "Yes." The other 99 responses indicated that the respondent had not been a CAA for Science test examiner, which ended the survey.

In the 2017–18 administration year, there were 3,140 active test examiners, indicating a 25 percent response rate from all available test examiners to the test examiner survey.

More than a quarter 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.

<u>Table 7.8</u> through <u>Table 7.16</u> provide the results for the test examiner surveys over all grade levels.

## 7.4.1. Responses to Background Questions

<u>Table 7.8</u> displays responses to the question, "Are you the teacher for the students you tested this year?" The data shows that just over 80 percent of test examiners were also the teacher for all of the students they tested.

Are you the teacher for the students you tested this year?	N=705 Test Examiners
a. Yes, I am the teacher for all of the students I tested this year.	81.3%
b. Yes, I am the teacher for some of the students I tested this year.	10.5%
c. No, I am not the teacher for any of the students I tested this year.	8.2%

#### Table 7.8 Teacher for Student Tested

Table 7.9 displays responses to the question, "To how many students did you administer the 2017–18 CAA for Science?" The data shows that more than half of test examiners administered the test to one to five students. More than a quarter of respondents indicated that they administered the test to 16 or more students.

#### Table 7.9 How Many Students Tested Per Test Examiner

To how many students did you administer the 2017–18 CAA for Science?	N=720 Test Examiners
a. 1–5 students	52.8%
b. 6–10 students	14.7%
c. 11–15 students	5.4%
d. 16 or more students	27.1%

<u>Table 7.10</u> displays the distribution of grades for which test examiners administered the CAA for Science in response to the question, "In what grade(s) did you administer the PTs? (Select all that apply.)."

Table 7.10 Grades Administere	bd
In what grade(s) did you administer the PTs? (Select all that apply.)	Percentages of Respondents
Grade five	47.6%
Grade eight	28.7%
High school	28.7%

#### Table 7.10 Grades Administered

## 7.4.2. Responses Regarding Testing Time

A majority of respondents indicated that they administered all three embedded PTs during a two-week period. <u>Table 7.11</u> displays test examiner responses to the question, "What approach did you take for administering all three PTs to the majority of students you tested?"

What approach did you take for administering all three performance tasks (PTs) to the majority of students you tested?	N=643 Test Examiners
a. The students were administered all three PTs during a two-week period.	80.4%
b. The students were administered all three PTs over several weeks.	16.6%
c. The students were administered all three PTs over several months.	3.0%

Question five of the survey was a "slider question," in which respondents interacted with a slider to indicate the cumulative time it took to administer all three PTs to a student. The most common responses, which are displayed in <u>Table 7.12</u>, were one hour, two hours, and three hours, at 26.1 percent, 26.1 percent, and 20.7 percent respectively.

#### Table 7.12 Cumulative Testing Time for All Three PTs Slider Question

On average, how much time (cumulative) did it take you to administer all three PTs to a student?	N=624 Test Examiners
a. 0 hours	1.9%
b. 1 hour	26.1%
c. 2 hours	26.1%
d. 3 hours	20.7%
e. 4 hours	9.3%
f. 5 hours	7.4%
g. 6 or more hours	8.5%

## 7.4.3. Responses Regarding Individualization

Test examiners were asked about how frequently they individualized the test for their students. Seventy-one percent indicated no individualization. All responses are displayed in <u>Table 7.13</u>.

How frequently did you take advantage of the option to "individualize" certain parts of the embedded PT, such as revised <i>Directions for Administration</i> (DFA) scripts or other materials for the activity?	N=572 Test Examiners
a. For all of my students, I individualized the test.	14.5%
b. For more than half of my students, I individualized the test.	1.2%
c. For around half of my students, I individualized the test.	3.1%
d. For fewer than half of my students, I individualized the test.	10.1%
e. For none of my students, I individualized the test.	71.0%

Table 7.13 Frequency of Individualization

The next question, asking about the importance of the individualizations, only appeared to the respondent if he or she indicated individualizing an embedded PT for some or all of students. As displayed in <u>Table 7.14</u>, more than 70 percent of test examiners indicated that these individualizations were important or very important to their student's performance on the assessment.

How important were the individualizations to your student's performance on the assessment?	N=133 Test Examiners
a. Very important	42.9%
b. Important	29.3%
c. Minimally important	21.8%
d. Not at all important	6.0%

Table 7.14 Importance of Individualization to Student Performance

Test examiners were also asked which elements of the task they individualized. Respondents only received this question for those PT(s) they had indicated administering in a previous question.

Results show that, depending on the PT, a range of 10.5 percent to 24.6 percent of test examiners administered the activity in a group setting. Depending on the PT, a range of 30.6 percent to 51.5 percent of test examiners individualized the script for their student. Finally, depending on the PT, a range of 31.6 percent to 53.9 percent of test examiners individualized the materials for their student. These results are displayed in <u>Table 7.15</u>.

Which elements of the task did you individualize?	Script	Materials	Administered Activity in Group Setting.	N=432 Test Examiners
Grade Five PT #1—"Sun and Shadows"	49.1%	31.6%	19.3%	57
Grade Five PT #2—"Physical Changes"	37.3%	42.4%	20.3%	59
Grade Five PT #3—"Weather Conditions"	43.9%	31.6%	24.6%	57
Grade Eight PT #1—"The Water Cycle"	50.0%	39.5%	10.5%	38
Grade Eight PT #2—"Bioenergy"	47.1%	35.3%	17.6%	34
Grade Eight PT #3—"Cells"	51.5%	33.3%	15.2%	33
High School PT #1—"Molecules"	30.8%	48.1%	21.2%	52
High School PT #2—"Force and Motion"	32.1%	50.9%	17.0%	53
High School PT #3—"Erosion"	30.6%	53.1%	16.3%	49

#### Table 7.15 Elements Individualized

## 7.4.4. Responses Regarding Clarity of Instructions Provided

Approximately 85 percent of respondents indicated that the directions for administration of the embedded PTs were either clear or reasonably clear. The results are displayed in <u>Table 7.16</u>.

How clear were the directions for administration in the embedded PT document?	N=504 Test Examiners
Clear	39.7%
Clear Reasonably clear	39.7% 44.6%

#### Table 7.16 Clarity of DFA

Respondents were also asked a short-answer question as to what aspects of the directions could be improved. Approximately 50 responses were received and were very diverse. Generally, respondents noted that the directions could be improved with more thought to the flow of testing, simpler step-by-step instructions, a reduction in the amount of prework to be done with paper, and the availability of a video that showed a teacher administering the test.

### 7.4.5. Open-ended Responses

Respondents were also asked a short-answer question about which aspects of the CAA for Science they liked. Approximately 330 diverse responses were received. Common responses included the availability of clear and engaging graphics, the hands-on aspect of

the test to promote learning, the option to substitute materials with concrete objects, how an embedded PT tested one concept at a time, a simplified testing process as compared with the previous pilot year, and the brevity of the test.

Respondents were also asked a short-answer question about what aspects of the CAA for Science could be improved. Approximately 330 diverse responses were received here as well. Common responses included the process of preparing materials and entering data into the Data Entry Interface, the difficulty of obtaining certain materials like a blow dryer and scale, the timing of access to directions (many would have preferred earlier access), the layout of the images, and the length of the test. A few responses indicated difficulty for students with moderate to severe cognitive disabilities and nonverbal students in answering certain items that were presented verbally, with no graphics to represent answer options.

## Chapter 8: 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 2017–18 California Alternate Assessment (CAA) for Science second-year pilot. The results of this analysis are summarized in this chapter. See also subsection <u>4.5 Accessibility Features for the Second-Year Pilot</u>, where individualizations are further described.

## 8.1. Considerations

The second-year pilot of the CAA for Science, which occurred during the 2017–18 California Assessment of Student Performance and Progress administration, provided Educational Testing Service (ETS) the first opportunity to collect data about this test to inform psychometric decisions.

One unique aspect of the CAA for Science second-year pilot 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, ETS psychometricians were concerned 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 activities? To answer this question, ETS conducted an evaluation to determine the impact on student performance by the decision to individualize and by material choices.

The test examiners downloaded the embedded PTs documentation which contained descriptions of the hands-on activity for each embedded PT (California Department of Education, 2018). The embedded PT documentation 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. The high school CAA for Science allowed for more individualization and flexibility than the grades five and eight CAAs for Science.

During the test administration, the test examiner used the Answer Recording Document to note the use of alternative materials and scripts, the amount of student engagement (fully engaged, moderately engaged, or minimally engaged), and the student's score for each test question. The test examiner then entered the information from the Answer Recording Document into the Data Entry Interface.

For the material choices analysis, the ETS Psychometrics and Research team used the statistical analysis sample, as described in subsection <u>6.1 Sample Used for the Analyses</u>, to examine the relationship between the choice of materials used and activity scores. The dependent variable for each of the models was the activity score, which is the sum of all points earned on items associated with each activity. The activity scores were analyzed because the questions regarding individualizations were asked at the activity level rather than the embedded PT level.

Embedded Performance Task and Test Comparability Considerations | Considerations

The independent variables of interest were:

- 1. **Disability type**—This variable indicated 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 activity (e.g., directions) were individualized in order to make the content accessible to the student. Its measurement used a coded variable (1 = yes; 0 = no).
- 3. **Student engagement**—This variable is a test examiner–reported measure completed after the administration of each activity, 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 activity. Its measurement used a coded variable (1 = yes; 0 = no).

These investigations were conducted only on groups of at least 25 students who used the same materials for the activity. For example, in grade five, only nine students had a visual impairment; these students were excluded from the analysis.

Several linear models were estimated for each embedded PT to evaluate the incremental impact of the test examiners' material choices had on the activity scores. 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 five linear models were estimated using SAS<sup>®</sup> PROC GLM module (general linear model). <u>Table 8.1</u> 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.

Model	Dependent Variable	Independent Variables
1	Activity score	<ul> <li>Disability</li> </ul>
2	Activity score	<ul> <li>Disability</li> </ul>
		<ul> <li>Individualization</li> </ul>
3	Activity score	<ul> <li>Disability</li> </ul>
		<ul> <li>Individualization</li> </ul>
		<ul> <li>Engagement</li> </ul>
4	Activity score	<ul> <li>Disability</li> </ul>
		<ul> <li>Individualization</li> </ul>
		<ul> <li>Engagement</li> </ul>
		<ul> <li>Materials</li> </ul>

Embedded Performance Task and Test Comparability Considerations | Summary of Findings from the Choice of Materials and Individualization Analysis

Model	Dependent Variable	Independent Variables
5	Activity score	Disability
		<ul> <li>Individualization</li> </ul>
		<ul> <li>Engagement</li> </ul>
		Materials
		<ul> <li>Interaction</li> </ul>

For each of the estimated models, the coefficient of determination, or R-squared ( $R^2$ ), was calculated. In addition, the change in  $R^2$  ( $R^2$  increment) was calculated to compare the differences in the increasingly complex statistical models. The  $R^2$  increment provides a convenient way to summarize the additional proportion of variance in activity scores explained by adding each independent variable into the model.

# 8.2. Summary of Findings from the Choice of Materials and Individualization Analysis

## 8.2.1. Individualization Analysis

The number and percentage of students in the statistical analysis sample using each of the individualizations is provided in <u>appendix 8.A</u>. In these tables, "NA" indicates that the question was not asked; often, the hands-on activity did not require the use of diagrams or pictures or materials. As an example, there were no questions regarding the use of individualized materials associated with PT 1 (Sun and Shadows), Activity 1, for grade five.

Fewer than five percent of the students included in the analyses for this investigation received an individualized script. While only three percent of the grade five students received an individualized diagram, more than eight percent of the grade eight students received an individualized diagram or picture. Additionally, approximately 20 percent of the students received individualized materials for the few embedded PTs that required materials for the hands-on activity.

## 8.2.2. Model Analysis

The linear models were run to analyze the relationship between the choice of materials and the students' activity scores. Summaries of the analyzed models are provided in <u>appendix 8.B</u>. The information included in these tables are as follows:

- Model: The variables included in the model
- R<sup>2</sup>: The proportion of the variance of the activity score explained by the independent variables specified in the model
- Difference in R<sup>2</sup>: The difference in R<sup>2</sup> 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

Embedded Performance Task and Test Comparability Considerations | Summary of Findings from the Choice of Materials and Individualization Analysis

- Partial eta-square (η<sup>2</sup>): 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 activities in which there was no choice of materials, the analyses included disability, individualization (e.g., individualized script, individualized diagram or picture), and student engagement as predictors. In these cases, only three models were included in the analysis (refer to <u>Table 8.B.1</u> as an example).

## 8.2.3. Results of the Individualization and Model Analyses

Overall, the proportion of variance in the students' activity scores accounted for by the models was low to moderate, with the R<sup>2</sup> ranging from 0.0289 to 0.2873. The value of R<sup>2</sup> increased minimally by including material choice to the model; the increase in R<sup>2</sup> ranged from 0.0002 to 0.0011. Therefore, adding material choice had negligible impact on the models. In the final models conducted for each activity, material choice accounted for a small amount of the variance. The partial  $\eta^2$  for material choice ranged between 0.0001 and 0.0014.

Adding student engagement to the models increased the value of R<sup>2</sup> and the amount of variance accounted for in the students' activity scores; the increase in R<sup>2</sup> ranged between 0.1255 and 0.2358. Additionally, for the final model run for each activity, student engagement accounted for more variance in the activity scores than any other variable included in the model. In the final models conducted for each activity, the partial  $\eta^2$  for student engagement ranged from 0.0095 to 0.2436. Partial  $\eta^2$  values of 0.14 or greater indicate that the student engagement has a large effect on the student's scores.

For grade five, only one activity included choice of materials for the hands-on, grade five PT 2 (Physical Changes) Activity 1. The results of the models conducted are provided in Table 8.B.2. The final model performed for this activity included disability, script, student engagement, and material choice. The choice to use individualized materials was not significant, and the partial  $\eta^2$  for materials was 0.0002. Therefore, the choice to use individualized materials does not explain a significant proportion of the variance in the students' activity score, given the other variables included in the model.

For high school, three activities allowed for choices of materials for the hands-on activity. These were PT 2 (Force and Motion) Activity 1, PT 2 (Force and Motion) Activity 2, and PT 3 (Erosion) Activity 1. To minimize any barriers to individualization that might be presented if materials were difficult to obtain, exemplar activities were designed to utilize either common classroom materials or materials considered to be easy for the test examiner to obtain. Some examples are books, small rubber balls, and cornmeal. Suggested substitute materials for individualization needs were selected based on this same criteria.

The physical science PTs had higher rates of individualizations because these activities allowed for more variation with regards to the material choices that test examiners could select. The results of these analyses are presented in <u>Table 8.B.12</u>, <u>Table 8.B.13</u>, and <u>Table 8.B.14</u>. The choice of using individualized materials is significant for PT 2 (Force and Motion) Activity 1. The interactions among the variables was modeled for PT 2 (Force and

Motion) Activity 1; the only factor that is significant is the interaction between disability and engagement.

# 8.3. Implications for the Field Test

### 8.3.1. 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 second-year pilot embedded PTs. Across all grades and activities, less than five percent of the students received an individualized script. At grades five and eight, less than 10 percent of the students received an individualized diagram or picture. When choices of materials were available at grade five and high school, only 19 to 22 percent of the students received individualized materials for the hands-on activity.

Second, in general, individualizations did not explain a significant proportion of the variance of the students' activity score, given the other variables in the model. For all the analyses, student engagement and disability explained significant proportions of the activity score, given the other variables in the model. Student engagement explained more variance in the activity scores, which is reflected by the partial  $\eta^2$ ; student engagement had larger partial  $\eta^2$  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  $R^2$  by including material choice in the model and in the small values of partial  $\eta^2$ .

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, individualized diagram or picture) or who received individualized materials. Additionally, the choice to use individualizations or individualized materials was made by the test examiner in order to make the hands-on 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.

### 8.3.2. Recommendations

For the 2018–19 CAA for Science field test administration, the ETS psychometric team recommends that information on individualizations—use of individualized scripts, diagrams, pictures, and materials—be collected from the test examiner. 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 content experts who are familiar with the CAA for Science student population review the field-test 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 California Next Generation Science Standards Core Content Connector being assessed by the test questions.

Embedded Performance Task and Test Comparability Considerations | References

## References

California Department of Education. (2018). CAA for Science 2018 Sample Task: Grade 5 embedded performance task, "Fossils." Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].

Cohen, J. (1988). *Statistical power analysis for behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

# Appendix 8.A: Choice of Materials and Individualization Analysis Data

**Note:** The data in <u>Table 8.A.1</u> through <u>Table 8.A.3</u> is based on the survey and includes only those students in the statistical analysis sample.

Individualization	PT 1 (Sun and Shadows), Activity 1—Number	PT 1 (Sun and Shadows), Activity 1—Percent of Total	PT 2 Physical Changes), Activity 1—Number	PT 2 (Physical Changes), Activity 1—Percent of Total	PT 3 (Weather Conditions), Activity 2—Number	PT 3 (Weather Conditions), Activity 2—Percent of Total
Using Standardized Scripts	3,747	96%	3,745	96%	3,817	98%
Using Individualized Scripts	115	3%	118	3%	84	2%
Using Standardized Diagram	3,775	96%	NA	NA	NA	NA
Using Individualized Diagram	115	3%	NA	NA	NA	NA
Using Standardized Materials	NA	NA	3,042	78%	NA	NA
Using Individualized Materials	NA	NA	840	22%	NA	NA

Table 8.A.1	Individualizations—Grade Five
-------------	-------------------------------

Table 8.A.2 Individualizations—Grade Eight (Bioenergy), Activity 2, Activity 1—Percent of Total of Total Activity 1—Percent of Total PT 2 (Bioenergy), Activity 2—Number 3 (Cells ), Activity 1-3 (Cells), Activity 2-PT 3 (Cells), Activity 2-PT 3 (Cells), Activity Percent of Total -Number Activity 2—Percent PT 2 (Bioenergy), Activity 1—Number Activity 2—Number PT 1 (Water Cycle), (Water Cycle), PT 1 (Water Cycle), PT 1 (Water Cycle), 2 (Bioenergy ), PT 2 (Bioenergy) Percent of Total Percent of Total Activity 1– PT 3 (Cel Number Number PT 1 F F Individualization 98% 3,786 Using Standardized Scripts 3,781 3,788 98% NA NA 98% 3,772 98% 3,784 98% Using Individualized Scripts 65 2% 60 2% NA NA 66 2% 84 2% 68 2% Using Standardized Diagram/ 3,495 90% 3,543 92% 3,480 NA NA 90% NA NA NA NA Picture 314 Using Individualized Diagram/ 357 9% 8% 373 10% NA NA NA NA NA NA Picture **Using Standardized Materials** NA **Using Individualized Materials** NA NA

						_				
Individualization	PT 1 (Molecules), Activity 1 Number	PT 1 (Molecules), Activity 1 Percent of Total	PT 1 (Molecules), Activity 2 Number	PT 1 (Molecules), Activity 2 Percent of Total	PT 2 (Force and Motion), Activity 1—Number	PT 2 (Force and Motion), Activity 1—Percent of Total	PT 2 (Force and Motion), Activity 2—Number	PT 2 (Force and Motion), Activity 2—Percent of Total	PT 3 (Erosion), Activity 1— Number	PT 3 (Erosion), Activity 1— Percent of Total
Using Standardized Scripts	5,348	96%	5,382	97%	5,412	97%	NA	NA	5,388	97%
Using Individualized Scripts	187	4%	156	3%	127	2%	NA	NA	146	3%
Using Standardized Diagram	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Using Individualized Diagram	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Using Standardized Materials	NA	NA	NA	NA	4,514	81%	4,452	80%	4,439	80%
Using Individualized Materials	NA	NA	NA	NA	1,033	19%	1,087	20%	1,055	19%

Table 8.A.3 Individualizations—High School

# Appendix 8.B: Model Analysis Summaries

	Table 8.B.1 Model Summary—Grade Five, PT 1, Activity 1								
Difference									
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>				
1	Disability	0.0735	NA	Disability: F,7 = 43.60 (<0.0001)	Disability: 0.0735				
2	Disability + Individualization	0.0754	0.0019	Disability: F,7 = 41.66 (<0.0001)	Disability: 0.0717				
				Script: F,1 = 9.22 (0.0024)	Script: 0.0024				
				Diagram: F,1 = 0.70 (0.4034)	Diagram: 0.0002				
3	Disability + Individualization +	0.2188	0.1434	Disability: F,7 = 33.26 (<0.0001)	Disability: 0.0582				
	Engagement			Script: F,1 = 8.55 (0.0035)	Script: 0.0023				
				Diagram: F,1 = 0.63 (0.4286)	Diagram: 0.0002				
				Engagement: F,2 = 345.99 (<0.0001)	Engagement: 0.1550				

Table 8.B.2 Model Summary—Grade Five, PT 2, Activity 1

			Difference		
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>
1	Disability	0.0531	NA	Disability: F,7 = 30.82 (<0.0001)	Disability: 0.0531
2	Disability + Individualization	0.0527	-0.0004	Disability: F,7 = 30.20 (<0.0001)	Disability: 0.0526
				Script: F,1 = 0.20 (0.6587)	Script: 0.0001
3	Disability + Individualization +	0.1782	0.1255	Disability: F,7 = 16.46 (<0.0001)	Disability: 0.0296
	Engagement			Script: F,1 = 0.14 (0.7080)	Script: 0.0000
				Engagement: F,2 = 288.38 (<0.0001)	Engagement: 0.1324
4	Disability + Individualization +	0.1787	0.0005	Disability: F,7 = 16.37 (<0.0001)	Disability: 0.0295
	Engagement + Materials			Script: F,1 = 0.35 (0.5529)	Script: 0.0001
				Engagement: F,2 = 287.07 (<0.0001)	Engagement: 0.1323
				Materials: F,1 = 0.91 (0.3412)	Materials: 0.0002

			-		
			Difference		
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>
1	Disability	0.0721	NA	Disability: F,7 = 42.67 (<0.0001)	Disability: 0.0721
2	Disability + Individualization	0.0721	0.000	Disability: F,7 = 42.31 (<0.0001)	Disability: 0.0719
				Script: F,1 = 0.10 (0.7528)	Script: 0.0000
3	Disability + Individualization +	0.2185	0.1464	Disability: F,7 = 27.38 (<0.0001)	Disability: 0.0478
	Engagement			Script: F,1 = 0.00 (0.9543)	Script: 0.0000
				Engagement: F,2 = 357.83 (<0.0001)	Engagement: 0.1577

### Table 8.B.3 Model Summary—Grade Five, PT 3, Activity 2

Table 8.B.4	Model Summary—G	Grade Eight, PT	1, Activity 1
-------------	-----------------	-----------------	---------------

	Difference								
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial n <sup>2</sup>				
1	Disability	0.0490	NA	Disability: F,8 = 24.63 (<0.0001)	Disability: 0.0490				
2	Disability + Individualization	0.0494	0.0004	Disability: F,8 = 24.42 (<0.0001)	Disability: 0.0493				
				Script: F,1 = 0.65 (0.4216)	Script: 0.0002				
				Diagram: F,1 = 1.07 (0.3012)	Diagram: 0.0003				
3	Disability + Individualization +	0.2032	0.1538	Disability: F,8 = 19.77 (<0.0001)	Disability: 0.0403				
	Engagement			Script: F,1 = 1.56 (0.2116)	Scrip: 0.0004				
				Diagram: F,1 = 0.09 (0.7705)	Diagram: 0.0000				
				Engagement: F,2 = 363.84 (<0.0001)	Engagement: 0.1618				

			•		
No.	Model	R <sup>2</sup>	Difference in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>
1	Disability	0.0474	NA	Disability: F,8 = 23.80 (<0.0001)	Disability: 0.0474
2	Disability + Individualization	0.0471	-0.0003	Disability: F,8 = 23.23 (<0.0001) Script: F,1 = 0.25 (0.6201)	Disability: 0.0469 Script: 0.0001
3	Disability + Individualization + Engagement	0.2171	0.1700	Diagram: F,1 = 0.07 (0.7924) Disability: F,8 = 15.70 (<0.0001) Script: F,1 = 0.14 (0.7036)	Diagram: 0.0000 Disability: 0.0322 Script: 0.0000
				Diagram: F,1 = 0.28 (0.5935) Engagement: F,2 = 410.12 (<0.0001)	Diagram: 0.0001 Engagement: 0.1784

#### Table 8.B.5 Model Summary—Grade Eight, PT 1, Activity 2

Table 8.B.6 Model Summary—Grade Eight, PT 2, Activity 1

	Difference							
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>			
1	Disability	0.0642	NA	Disability: F,8 = 32.80 (<0.0001)	Disability: 0.0642			
2	Disability + Individualization	0.0641	-0.0001	Disability: F,8 = 32.34 (<0.0001)	Disability: 0.0639			
_				Picture: F,1 = 0.24 (0.6262)	Picture: 0.0001			
3	Disability + Individualization +	0.2184	0.1543	Disability: F,8 = 20.14 (<0.0001)	Disability: 0.0408			
	Engagement			Picture: F,1 = 1.22 (0.2701)	Picture: 0.0003			
				Engagement: F,2 = 373.90 (<0.0001)	Engagement: 0.1649			

Table 8.B.7	Model Summary-	-Grade Eight,	, PT 2, Activity 2
-------------	----------------	---------------	--------------------

	Difference								
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>				
1	Disability	0.0672	NA	Disability: F,8 = 34.43 (<0.0001)	Disability: 0.0672				
2	Disability + Individualization	0.0664	-0.0008	Disability: F,8 = 33.64 (<0.0001) Script: F,1 = 0.41 (0.5222)	Disability: 0.0663 Script: 0.0001				
3	Disability + Individualization + Engagement	0.2381	0.1717	Disability: F,8 = 21.78 (<0.0001) Script: F,1 = 1.77 (0.1840) Engagement: F,2 = 426.81 (<0.0001)	Disability: 0.0440 Script: 0.0005 Engagement: 0.1840				

	Difference					
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>	
1	Disability	0.0663	NA	Disability: F,8 = 33.95 (<0.0001)	Disability: 0.0663	
2	Disability + Individualization	0.0683	0.0020	Disability: F,8 = 33.63 (<0.0001)	Disability: 0.0663	
				Script: F,1 = 5.61 (0.0179)	Script: 0.0015	
3	Disability + Individualization +	0.2873	0.2190	Disability: F,8 = 24.20 (<0.0001)	Disability: 0.0486	
	Engagement			Script: F,1 = 7.22 (0.0073)	Script: 0.0019	
				Engagement: F,2 = 581.77 (<0.0001)	Engagement: 0.2351	

#### Table 8.B.8 Model Summary—Grade Eight, PT 3, Activity 1

#### Table 8.B.9 Model Summary—Grade Eight, PT 3, Activity 2

	Difference					
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>	
1	Disability	0.0613	NA	Disability: F,8 = 31.22 (<0.0001)	Disability: 0.0613	
2	Disability + Individualization	0.0617	0.0004	Disability: F,8 = 31.10 (<0.0001)	Disability: 0.0617	
				Script: F,1 = 0.03 (0.8580)	Script: 0.0000	
3	Disability + Individualization +	0.2621	0.2004	Disability: F,8 = 18.18 (<0.0001)	Disability: 0.0370	
	Engagement			Script: F,1 = 0.00 (0.9789)	Script: 0.0000	
				Engagement: F,2 = 513.26 (<0.0001)	Engagement: 0.2135	

#### Table 8.B.10 Model Summary—High School, PT 1, Activity 1

	Difference				
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>
1	Disability	0.0317	NA	Disability: F,10 = 18.18 (<0.0001)	Disability: 0.0317
2	Disability + Individualization	0.0328	0.0011	Disability: F,10 = 18.07 (<0.0001)	Disability: 0.0317
				Script: F,1 = 5.86 (0.0155)	Script: 0.0011
3	Disability + Individualization +	0.1986	0.1658	Disability: F,10 = 13.60 (<0.0001)	Disability: 0.0241
	Engagement			Script: F,1 = 2.91 (0.0880)	Script: 0.0005
				Engagement: F,2 = 569.68 (<0.0001)	Engagement: 0.1717

	Difference					
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>	
1	Disability	0.0297	NA	Disability: F,10 = 16.97 (<0.0001)	Disability: 0.0297	
2	Disability + Individualization	0.0301	0.0004	Disability: F,10 = 16.92 (<0.0001)	Disability: 0.0297	
				Script: F,1 = 1.53 (0.2163)	Script: 0.0003	
3	Disability + Individualization +	0.2370	0.2069	Disability: F,10 = 13.61 (<0.0001)	Disability: 0.0241	
	Engagement			Script: F,1 = 0.30 (0.5838)	Script: 0.0001	
				Engagement: F,2 = 745.82 (<0.0001)	Engagement: 0.2133	

### Table 8.B.11 Model Summary—High School, PT 1, Activity 2

	Difference					
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial η <sup>2</sup>	
1	Disability	0.0289	NA	Disability: F,10 = 16.49 (<0.0001)	Disability: 0.0289	
2	Disability + Individualization	0.0295	0.0006	Disability: F,10 = 16.22 (<0.0001)	Disability: 0.0285	
				Script: F,1 = 4.77 (0.0290)	Script: 0.0009	
3	Disability + Individualization +	0.2587	0.2292	Disability: F,10 = 10.84 (<0.0001)	Disability: 0.0194	
	Engagement			Script: F,1 = 2.21 (0.1372)	Script: 0.0004	
				Engagement: F,2 = 846.99 (<0.0001)	Engagement: 0.2360	
4	Disability + Individualization +	0.2597	0.0010	Disability: F,10 = 10.63 (<0.0001)	Disability: 0.0191	
	Engagement + Materials			Script: F,1 = 4.00 (0.0456)	Script: 0.0007	
				Engagement: F,2 = 841.59 (<0.0001)	Engagement: 0.2353	
				Material: F,1 = 12.20 (0.0005)	Material: 0.0022	
5	Disability + Individualization +	0.2708	0.0111	Interaction of Disability by	Disability: 0.051	
	Engagement + Materials with			Engagement: F,20 = 1.60 (0.0447)	Script: 0.0003	
	Interactions				Engagement: 0.0095	
					Material: 0.0001	
					Interaction of	
					Disability by	
					Engagement: 0.0059	

Difference R<sup>2</sup> Model in R<sup>2</sup> No. Significance Tests Partial  $\eta^2$ Disability 0.0319 NA Disability: F, 10 = 18.30 (<0.0001) Disability: 0.0319 1 Disability: F, 10 = 13.21 (<0.0001) 2 Disability + Engagement 0.2358 Disability: 0.0235 0.2677 Engagement: F,2 = 882.91 (<0.0001) Engagement: 0.2436 3 Disability + Engagement + 0.2688 0.0011 Disability: F, 10 = 13.03 (<0.0001) Disability: 0.0232 Materials Engagement: 0.2436 Engagement:  $F_{,2} = 882.57 (< 0.0001)$ Materials: F, 1 = 7.84 (0.0051) Materials: 0.0014

#### Table 8.B.13 Model Summary—High School, PT 2, Activity 2

Table 8.B.14 Model Summary—High School	ol, PT 3, Activity 1
--	----------------------

	Difference					
No.	Model	R <sup>2</sup>	in R <sup>2</sup>	Significance Tests	Partial n <sup>2</sup>	
1	Disability	0.0436	NA	Disability: F,10 = 25.28 (<0.0001)	Disability: 0.0436	
2	Disability + Individualization	0.0445	0.0009	Disability: F,10 = 25.32 (<0.0001)	Disability: 0.0443	
				Script: F, 1 = 0.24 (0.6236)	Script: 0.0000	
3	Disability + Individualization +	0.2539	0.2094	Disability: F,10 = 18.46 (<0.0001)	Disability: 0.0325	
	Engagement			Script: F, 1 = 0.11 (0.7369)	Script: 0.0000	
				Engagement: F,2 = 772.66 (<0.0001)	Engagement: 0.2196	
4	Disability + Individualization +	0.2541	0.0002	Disability: F,10 = 18.69 (<0.0001)	Disability: 0.0333	
	Engagement + Materials			Script: F, 1 = 0.06 (0.8111)	Script: 0.0000	
				Engagement: F,2 = 761.22 (<0.0001)	Engagement: 0.2189	
				Materials: $F,1 = 0.31$ (0.5784)	Materials: 0.0001	

# **Chapter 9: 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 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 of evaluating those products and services. Quality control procedures are outlined in this chapter.

# 9.1. 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 undergo an extensive embedded PT review process. The item writers hired to develop CAA items were trained in California Assessment of Student Performance and Progress (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 is accepted for authoring—that is, once it has been entered into ETS' item bank and formatted for use in an assessment—ETS employs a series of internal and external reviews. These reviews use established criteria and specifications to judge the quality of item content and to ensure that each item measures what it is intended to measure. These reviews also examine the overall quality of the test items before presentation to the CDE and item reviewers. Finally, a group of California educators review the items for accessibility, bias/sensitivity, and content prior to their administration to students. The details on quality control of item development are described in subsection <u>3.2 Embedded PT Review Process</u>.

# 9.2. 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. Although the blueprint for the CAA for Science was not finalized prior to the test form assembly, the CDE and ETS had many conversations concerning the format of assembly and delivery. The mutually agreed upon design was implemented for this the CAA for Science second-year pilot.

### 9.2.1. Quality Control of Test Form Development

ETS conducts multiple levels of quality assurance checks on each constructed test form to ensure it meets defined statistical criteria. These quality assurance checks are critical to overall test integrity. For the 2017–18 CAA for Science, both ETS assessment development and psychometric staff reviewed and signed off on the accuracy of forms before the test forms were put into production for the second-year pilot. Detailed information related to test assembly can be found in subsection <u>3.2.1 Selection of Embedded PTs</u>.

In particular, the assembly of all test forms went through a certification process that included various checks to verify that

- all correct answers are accurately designated,
- answers items are scored correctly in the item bank,
- all embedded PTs match the standard,
- all content in the embedded PT is correct,
- distractors are plausible,
- multiple-choice item options are parallel in structure,
- language is grade-level appropriate for this population,
- no more than three multiple-choice items in a row have the same key,
- all art is correct,
- there are no errors in spelling or grammar, and
- embedded PTs adhere to the approved style guide.

Reviews were also conducted for functionality and sequencing during the user acceptance testing process to ensure all items functioned as expected.

### 9.2.2. 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, 2018d). The two systems are kept in sync during the testing window. Students at eligible grade levels are 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 twelve—local educational agencies (LEAs) log on to TOMS and assign students to take the alternate assessment, which automatically unassigns those students from taking the CAST. Additionally, should the LEA determine that a student in grade ten or eleven will take the high school CAA for Science, the LEA will assign that student to test.

The quality of test assignment for the CAA for Science is monitored and controlled through several strategies. TOMS enforces preconditions for eligibility for the CAAs by permitting assignment only for students with an Individuals with Disabilities Education Act<sup>5</sup> 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 (CALPADS Field 3.21) is not set to blank.

Additionally, TOMS prevents 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.

<sup>&</sup>lt;sup>5</sup> 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.

### 9.2.3. 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 receive training on these topics and are provided with tools to report security incidents and resolve testing discrepancies for specific testing sessions.

As is true for all assessments administered as part of the CAASPP System, several strategies are utilized to monitor and control the quality of test administration for the CAA for Science. A fully staffed support center, the California Technical Assistance Center (CaITAC), supports all LEAs in the administration of CAASPP assessments. CaITAC is guided by a core group of LEA outreach and advocacy staff who manage communications to LEAs, regional and web-based trainings, and a website, <a href="http://www.caaspp.org/">http://www.caaspp.org/</a>, that houses a full range of manuals, videos, and other instructional and support materials. In addition to providing guidance and answering questions, CaITAC 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. OTI's detailed quality-control procedures are described in subsection <u>4.6.1. ETS'</u> <u>Office of Testing Integrity (OTI)</u>.

# 9.3. Quality Control of Test Materials

The steps taken to develop and ensure the quality of the online assessments are described in <u>Chapter 3: Embedded Performance Task Development and Review</u>.

### 9.3.1. Test Administration Manuals

ETS staff consult with internal subject matter experts and conduct validation checks to verify that test directions and administration manuals accurately match the test materials and testing processes. Copy editors and content editors review each document for spelling, grammar, accuracy, and adherence to CDE style. Each document must be approved by the CDE before it can be published to the CAASPP Portal at <a href="http://www.caaspp.org/">http://www.caaspp.org/</a>. Only nonsecure documents are posted to this website. Secure materials, such as the CAA for Science Embedded Performance Task Directions for Administration, are made available to designated LEA staff through TOMS, which requires a secure log on.

The manuals used in the administration of the CAA are listed in subsection <u>4.4.4 Instructions for Test Examiners and Staff Involved in CAA Administration</u>.

### 9.3.2. Processing Test Materials

The following information was transcribed on the 2017–18 CAA for Science answer recording sheet, entered into the Data Entry Interface (DEI) by test examiners, and transmitted from the American Institutes for Research (AIR) to ETS through the DEI each day:

- Student's first name
- Statewide Student Identifier
- Results of the Student Response Check
- Any individualized scripts and materials used
- Scores 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.

# 9.4. Quality Control of Psychometric Processes

### 9.4.1. Development of Scoring Specifications

ETS scoring specifications for the CAA for Science are completed, approved, and checked well in advance of the receipt of student response data. These specifications contain detailed scoring procedures, as well as the procedures for determining whether a student has attempted a test and whether that student's response data should be included in the statistical analyses and calculations for computing summary data.

### 9.4.2. Development of Scoring Procedures

Following scoring, a series of quality control checks are carried out by ETS psychometricians to ensure the accuracy of each score.

### 9.4.3. Psychometric Processing

The psychometric analyses conducted at ETS undergo comprehensive quality checks by a team of psychometricians and data analysts. The ETS psychometric team reviews the data files before conducting the statistical analyses to ensure the quality of the data. The team develops 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 are run by one data analyst and checked by a second data analyst. Results are then reviewed by the psychometricians to compile a list of flagged items for ETS Assessment Development (AD) staff for review. AD comments are reviewed by the psychometricians before items are approved for inclusion in additional analyses and before the data review meetings with the CDE.

# 9.5. Quality Control of Reporting

The CAA for Science second-year pilot tests were scored at the local level by the test examiner or LEA staff. ETS provided instructions to test examiners in the *Embedded Performance Task Direction for Administration* (CDE, 2018a). The test examiner or LEA staff computed the overall percent correct scores (described in subsection <u>5.2.1 Percent Correct</u>) and preliminary indicator category (described in subsection <u>5.2.2 Preliminary Indicator Categories</u>) for each student. ETS also provided a scoring tool to help the LEAs calculate the student's overall score, percent correct, and preliminary indicator category. More information about the preliminary indicators can be found on the CDE Preliminary Indicator Communication Toolkit web page (CDE, 2018c).

The ETS psychometric team checked the scoring tool by simulating all possible raw score combinations across the three PTs and computing the total raw score, the percent correct, and preliminary indicator category. All possible scenarios of incompletion of the test (e.g., student completes two of the three PTs) were also simulated as part of the quality control process. The psychometric team computed the scores using both the scoring tool and independently using SAS<sup>®</sup> and then compared the results. Any differences in results were investigated and issues were resolved. This process was repeated until the results from scoring tool and the independent calculations matched.

An aggregate report summarizing the results of the 2017–18 second-year pilot administration for the CAA for Science was provided to the CDE. CAA for Science scores for a given grade are aggregated and generated at the school, LEA or direct funded charter school, county, and state levels. To ensure the quality of the aggregate report, two members of the ETS psychometric team individually produced an aggregate file, including the percent correct and preliminary indicator category for each student aggregated at group levels. The files and student and group results were compared and any differences were resolved. A third member of the psychometric team reviewed the aggregate file for reasonableness and spot checked the numbers for accuracy.

## References

- California Department of Education. (2018a). CAA for Science 2018 Sample Task: Grade 5 embedded performance task, "Fossils." Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].
- California Department of Education. (2018b). *California Alternate Assessment for Science blueprint*. California Department of Education. Retrieved from <a href="https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx">https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx</a>
- California Department of Education. (2018c). *Preliminary indicator communication toolkit.* Retrieved from [This link is no longer active]

California Department of Education. (2018d). TOMS pre-administration guide for CAASPP testing. Sacramento, CA: California Department of Education. Retrieved from [this link is no longer active].

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

Educational Testing Service. Retrieved from <u>https://www.ets.org/s/about/pdf/</u> <u>standards.pdf</u>

# Chapter 10: Continuous and Systematic Improvements

# **10.1.** Improvements from the First-Year Pilot

There were several changes made to the embedded performance task (PT) format for the second-year pilot administration of the California Alternate Assessment (CAA) for Science based on the lessons learned from the first-year pilot.

### 10.1.1. Changes to Test Administration

The availability of feedback in the form of the preliminary indicators was a major change to the administration of the second-year pilot embedded PTs in comparison with those of the first-year pilot. For the second-year pilot administration, test examiners were tasked with recording student responses for later entry into American Institutes for Research's Data Entry Interface (DEI).

#### 10.1.1.1 Student Response

A new feature of the second-year pilot was the addition of a Student Response Check (SRC) to provide clear guidance in each task about the conditions in which testing should stop if a student did not respond. Prior to administering the embedded PT, test examiners were directed to conduct an SRC with the student. The purpose of the SRC is for the test examiner to verify if the student had a consistent and observable way of indicating a response to test questions. Test examiners were directed to administer this check by presenting any three objects specified in the materials list of the PT document and then directing the student to identify one familiar object in the set of objects using the student's mode of communication. If the student was able to orient and respond to the SRC, then the test examiner was directed to continue with the administration of the entire test. If the student did not orient and respond during the SRC, then the test examiner was directed to end test administration of that particular PT.

#### 10.1.1.2 Student Survey Questions

In both the first-year pilot and the second-year pilot embedded PTs, test examiners were directed to ask a short, two-question survey of students.

The first question asked how the student felt about taking the test, with answer options being happy, sad or confused. This question was accompanied by graphics displaying happy, sad, and confused faces.

The second question asked "Did you have enough time to complete the test?" and had a Yes or No option accompanied by graphics displaying a thumbs up and a thumbs down. These two survey questions were kept in order to continue to gather information about student's experiences taking the test. Refer to subsection <u>7.2 Student Survey Administration</u> for more information about the survey presentation. Refer to subsection <u>7.3 Student Survey Results</u> for survey results.

### **10.1.2. Answer Recording Document**

A consistent feedback received from test examiners regarding the first-year pilot pertained to the difficulty of knowing where to record student responses. Given that student results would be collected during the 2017–18 administration via the DEI, a formal Answer Recording Document that simplified the recording of student responses was provided in each embedded PT. During test administration, the test examiner would record the student

responses in the Answer Recording Document for later entry in the DEI. Throughout the PT *Directions for Administration (DFA)* document, a diamond shape with a numbered label flagged information that corresponded to a number on the Answer Recording Document where a data input would be needed. Thereafter, the Answer Recording Document would serve as the local educational agency's record of student results.

### 10.1.3. Student Independence Evaluations

The task of evaluating the independence of student responses at multiple points in the assessment was eliminated. This was replaced by a simpler evaluation completed at the end of the PT.

### 10.1.4. Changes to Content

Based on feedback from test examiners, the overall complexity of the questions was reduced. Individual items were shortened, and the number of items per task was reduced for most tasks. The language was also simplified and streamlined.

The complexity of hands-on activities was somewhat reduced. The number of lengthy hands-on activities was also reduced by using graphics-based activities more extensively. This reduced the amount of materials preparation required of examiners.

# 10.2. In-Person Observations

For the second-year pilot, in-person observations of students and test examiners were conducted during the administration of an embedded PT. The data from this study was used to collect evidence about the usability and understandability of test materials.

As part of the test development process, it is important to examine the performance of the test items when taken by students who represent the diversity of the intended population of test takers for the assessment. In addition to large-scale research efforts like pilot testing and field testing, which focus on item performance, small-scale research studies can provide a more in-depth and interactive collection of information. This information is particularly important to collect from students who may interact with or interpret an assessment differently, including students with a variety of disabilities, as well as from the test examiners who administer assessments individually.

Nine schools participated in this voluntary study. The students studied were eligible to take the 2017–18 CAA for Science. A parent consent form for each participating student and a principal consent form for each participating school were collected.

### 10.2.1. Observation Protocols

This study employed an observational research approach to ascertain the aspects of the embedded PTs that appeared to be understandable and used as expected, and to identify the features of the embedded PTs that were problematic. Due to the characteristics of the students who qualify to take the alternate assessment, this study was designed to capture data based on the student's interactions with the CAA for Science embedded PTs and collect additional feedback through the interview questions directed to test examiners. The factors that led to this design choice are as follows:

- Students eligible to take this test are identified as significantly cognitively disabled.
- Some students eligible to take this test are nonverbal.
- For this population, students' responses may vary due to factors other than the knowledge and skills being assessed. A student may respond with more or less focus

Continuous and Systematic Improvements | In-Person Observations

depending on other external conditions, which would be better known by a test examiner who is familiar with the students range of abilities, focus, and mode of response.

The protocol was designed by an educational assessment research scientist experienced in qualitative research and familiar with the population of students with disabilities. The protocol document was used to guide observations and interview questions with the goal of collecting objective information about what the students were doing and understanding; it allowed notes to be captured during and after the observations.

### 10.2.2. Summary of the In-Person Observations

Prior to data collection, the Educational Testing Service (ETS) team members responsible for conducting each of the study sessions participated in a four-hour online training. The training reviewed the purpose of the study, study materials, major sections of the protocol, and level of details expected in observation notes.

Thirty-eight students and 12 test examiners from nine schools participated in this study. <u>Table 10.1</u> shows the sample of participating students and test examiners by grade.

Table 10.1 Number of Student Participants and Test Examiners by Grade

Participant	Grade 5	Grade 8	Grade 11	Grade 12	Total
Student	14	7	5	12	38
Test Examiner	3	4	1	4	12

Table 10.2 shows the embedded PTs for which observations were collected.

Grade Level	Embedded PT	Observations (N)
5	Physical Changes	3
5	Sun and Shadows	4
5	Weather Conditions	7
8	Cells	6
8	The Water Cycle	1
High School	Force and Motion	13
High School	Erosion	1
High School	Molecules	3

Table 10.2 Number of Observations	by Grade and Embedded PT
-----------------------------------	--------------------------

Before the observational session started, the observer requested assent of the student and test examiner. Thereafter, observations were conducted as testing proceeded. Upon the completion of the student engagement survey, the student was dismissed from the study session and the ETS observer conducted the postobservational interview; the observer asked the text examiner the protocol questions and recorded the test examiner's responses.

### 10.2.3. Lessons Learned from the In-Person Observations

### 10.2.3.1 Administration by Test Examiners

Generally, test examiners set up materials prior to testing; ETS staff reviewed the tasks being set up and observed the SRC. Overall, the materials were being set up appropriately. Observers noted that most test examiners appeared to have read the *DFA* and were following the instructions.

Test examiners largely seemed to be familiar with the materials and followed the instructions in the *DFA*. All test examiners read the scripts. Some test examiners used some prompting, which included repeating the question, using hand gestures, and describing pictures. Most used the exemplar activity.

Test examiners were flexible to the needs of the students while also using materials that were available. For example, the exemplar for the high school embedded PT Force and Motion suggests using a small ball; one test examiner had a basketball with a cardboard box set up as a ramp. Some test examiners printed materials and graphics prior to testing a student. One test examiner went into the school garden to administer the high school–level Erosion task.

Test examiners accommodated students who, in addition to significant cognitive disabilities, had additional disabilities, such as students with visual impairments.

#### 10.2.3.2 Student Responses

Generally, students were observed responding by pointing, responding verbally, or both. Most appeared to understand the activity. Although some test examiners shared concerns about graphics, observations of student interaction with graphics were mixed. Some students did not focus on graphics, and others appeared to find images of mountains and jungles interesting.

### 10.2.4. Postobservation Interviews

Four themes emerged from the interviews:

- 1. Challenges regarding preparation and suggestions regarding test materials
- 2. Questions and suggestions about administration instructions
- 3. Teaching methods for the topics tested
- 4. Test examiner impressions of the student interactions, with overall comments and suggestions for the future

#### 10.2.4.1 Challenges Regarding Preparation and Suggestions Regarding Test Materials

Three themes emerged from test examiner comments regarding preparation and materials:

- 1. Preparation and availability of materials prior to the administration
- 2. The amount of test materials test examiners had to work with
- 3. The use of graphics

In discussions about preparing prior to the administration, multiple test examiners expressed a wish to practice with the materials ahead of time. Other test examiners who compared the use of materials in the first-year pilot with the second-year pilot found some level of improvement.

Comments noted by ETS observers include the following:

- "Would have been good to have more practice, mock trials, before giving the real test to students."
- "This year there was a lot of prep done for teachers, e.g., pre-cut materials. That helped with flow, (as did the) embedded instructions."
- "One test examiner stated that they had to spend more time preparing last year; easier this year than last, was harder overall last year, graphics were helpful."

Continuous and Systematic Improvements | Implications Based on the Survey Results

Another theme surrounding test materials was that there was too much. Printed documents made it challenging for the test examiner. There were too many papers with different purposes needed at different times, which impacted the student's ability to pay attention. Some suggestions were included in comments.

- "It was challenging to match up the various documents during testing."
- "Had to adapt the answers so that the student was able to point or adapt to the answer (with greater visual stimulation). These students do not typically read."

The third theme regarding test materials was about the use of graphics. The comments range from stating that graphics were helpful to stating that graphics needed color; however, some comments also stated that graphics may not be good for this population, the implication being that students in this population may benefit from tangible materials instead.

#### **10.2.4.2 Questions and Suggestions About Administration Instructions**

All test examiners reported that they understood when to administer the test and when to administer the SRC. A small number of test examiners had questions regarding the instructions on how to administer the test or how much flexibility the test examiners had.

#### 10.2.4.3 Teaching Methods of Instruction

Most test examiners did not comment on teaching method. Among the comments that were collected, most said they covered the topic with similar activities to the tasks presented in the assessment, such as,

- "class grew a basil plant,"
- "created a cell using construction paper,"
- "read stories about things that are alive and not alive," and
- "demonstrations, videos, direct instruction, a lot of visuals."

One test examiner reported that he or she teaches mostly functional life skills but is incorporating more science with the introduction of the CAA for Science.

#### **10.2.4.4 Overall Comments and Suggestions for the Future**

When test examiners were asked for overall comments or suggestions for future tasks, the theme that arose most often was about the mode of administration. Computer-based testing, video images, and animation were common suggestions. One test examiner who administered the assessment to three students mentioned that the school has a functional skills curriculum and suggested that, for these students, the test seems less appropriate. He cited vocabulary as an example of what is beyond the students' comprehension.

There were positive overall comments as well, such as, "I appreciate that we are working on something meaningful," and "I am glad we are scoring the test this year for Pilot 2, unlike Pilot 1 last year."

# **10.3. Implications Based on the Survey Results**

For the most part, the CAA for Science PTs were accessible to the students and the test examiners; the test materials and instructions were clear and understandable to most of the participants.

### 10.3.1. Test Delivery

Some themes that emerged included a need for a streamlined and more manageable approach to the many documents needed. Suggestions included online tests.

To address some of the concerns raised, the 2019–20 of the CAA for Science operational assessment will be delivered online, alleviating the need to manage test administration materials.

### 10.3.2. Use of Graphics

Another topic raised was the use of graphics. However, because the comments from test examiners included some suggestions that were specific to the student and educator in the study, there were some contradictions. Some stated that graphics may be too abstract, with the implication that they should be replaced. Others suggested that the graphics be modified by the use of color and made larger. These suggestions and concerns are complicated, which was pointed out by some test examiners who mentioned specific problems for students who are visually impaired.

The introduction of online testing may allow more flexibility for the graphics, videos, and animations that could be included in the future. However, given the specific disabilities that some students have in addition to severe cognitive disabilities, the solution may be complicated.

### 10.3.3. Disseminating Information About the CAA for Science

In the second-year pilot, test examiners saw some improvements from the assessment they administered the previous year; however, they expressed a continued need for more information to be shared. Specifically, test examiners made a few comments about opportunities to increase test examiners' familiarity with the embedded PTs—in discussions about preparing prior to the administration, multiple test examiners expressed a wish to practice with the materials ahead of time.

The communication needed includes when the embedded PTs can be administered, along with the recommendation that each PT can be administered soon after classroom instruction on the topic.

It is important to note the difference between being familiar with the materials and using the task as part of instruction. Test examiners should become familiar with the PT and materials while continuing to teach the topic as appropriate to their students and not practice the actual PT with the student so they will learn the task; doing so would not allow the actual test to measure the student's ability.

One important result in this study is that teachers indicated they are teaching these topics in ways that are similar to the materials being used in the assessment. This is a crucial part of the chain of validity evidence for any student group tested and may be more important in this specific population. For these students, where the mode of instruction may be individualized, the results here support the need for the CAA for Science to allow flexibility, so that the assessment is aligned with the students' specific needs, such as a visual impairment or variability in communication.