California Department of Education
Assessment Development & Administration Division

California Assessment of Student Performance and Progress California Alternate Assessment for Science 2019–2020 Technical Report

Submitted July 2, 2021

By ETS

Contract #CN150012

Table of Contents

List of Tables

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

Term	Definition
1PL	one-parameter logistic
1PL-IRT	one-parameter logistic item response theory
AERA	American Educational Research Association
AIS	average item score
ALTRD	Assessment and Learning Technology Research & Development
APA	American Psychological Association
CA NGSS	California Next Generation Science Standards
CAA	California Alternate Assessment
CAASPP	California Assessment of Student Performance and Progress
CAC	California Assessment Conference
CALPADS	California Longitudinal Pupil Achievement Data System
CAI	Cambium Assessment, Inc.
CalTAC	California Technical Assistance Center
CAST	California Science Test
CCR	*California Code of Regulations*
CCSS	Common Core State Standards
CDE	California Department of Education
CERS	California Educator Reporting System
COVID-19	novel coronavirus disease 2019
DFA	*Directions for Administration*
DIF	differential item functioning
EC	*Education Code*
ELA	English language arts/literacy
ELD	English Language Development
ELPAC	English Language Proficiency Assessments for California
eSKM	Enterprise Score Key Management
EUs	essential understandings
FKSA	focal knowledge, skills, and abilities
GPCM	generalized partial credit model
IEP	individualized education program
IRT	item response theory
ISAAP	Individual Student Assessment Accessibility Profile
JAWS	Job Access With Speech
LEA	local educational agency
MH	Mantel-Haenszel
NCME	National Council on Measurement in Education
ORS	Online Reporting System
OTI	Office of Testing Integrity

Table of Acronyms and Initialisms *(continuation)*

Term	Definition
PCM	partial credit model
PE	performance expectation
PT	performance task
QA	quality assurance
RMSEA	root mean square error of approximation
SBE	State Board of Education
Science Connectors	Science Core Content Connectors
SCOE	Sacramento County Office of Education
SD	standard deviation
SEM	standard error of measurement
SFTP	Secure File Transfer Protocol
SMD	standardized mean difference
SR	selected response
SSID	Statewide Student Identifier
STAIRS	Security and Test Administration Incident Reporting System
TDS	test delivery system
TEI	technology-enhanced items
TOMS	Test Operations Management System
UAT	user acceptance testing
USC	United States Code

Chapter 1: Introduction

This chapter provides an overview of the California Alternate Assessment (CAA) for Science program, including background information, the purpose of the test, the intended population, and organizations and systems involved. Additionally, this chapter provides a summary of the impact of the novel coronavirus disease 2019 (COVID-19) pandemic on psychometric analyses and reporting (refer to section *1.8. Impact of the Novel Coronavirus Disease 2019 Pandemic on Psychometric Analyses and Reporting*).

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 CAA for Science is an assessment aligned with the Science Core Content Connectors (Science Connectors) derived from the CA NGSS. Its field test was administered during the 2018–2019 CAASPP administration.

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

The CAA for Science is for students in grades five and eight and in high school whose individualized education program (IEP) teams have determined that alternate assessments are appropriate for the student (CDE, 2020a). 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 2019–2020 administration, the CAASPP System comprised the following assessments:

- Smarter Balanced assessments and tools for the general student population:
  - Summative Assessments—Online assessments for ELA and mathematics in grades three through eight and grade eleven
  - Interim Assessments—Optional resources developed for grades three through eight and grade eleven designed to inform and promote teaching and learning by providing information that can be used to monitor student progress toward mastery of the Common Core State Standards that may be administered to students at any grade level
  - Digital Library (now Tools for Teachers)—Professional development materials and instructional resources designed to help teachers use formative assessment processes for improved teaching and learning in all grades

- CAAs for ELA and mathematics in grades three through eight and grade eleven for students with significant cognitive disabilities
- Science assessments in grades five and eight and high school (grade ten, eleven, or twelve; these are the California Science Test [CAST] and the CAA for Science)
- The California Spanish Assessment, optional for eligible students in grades three through eight and high school and designed to measure a student's Spanish competency in reading, writing mechanics, and listening, as well as a high school measure suitable to be used in part for the California Seal of Biliteracy

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

1.2. Test Purpose

The purpose of the CAA for Science is to measure what students know and can do. These measures help identify and address gaps in knowledge or skills (CDE, 2020a). The CAA for Science assesses 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.

1.3. Content and Structure

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

1.3.1. Assessment Model

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

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

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

California's relatively small population of students with the most significant cognitive disabilities who are eligible for an alternate science assessment[1] also makes the use of this assessment model reasonable.

1.3.2. California Next Generation Science Standards Core Content Connectors

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

Table 1.1 Organization of the Science Connectors

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

1.3.3. Test Components for the 2019–2020 Administration

The 2019–2020 CAA for Science involved three components:

1. Four embedded PTs (Refer to section *5.3 Administration of the Embedded Performance Tasks* for a description of how these were administered.)
2. A brief survey to collect information about the student's responsiveness to the embedded performance task (Refer to *Chapter 8: Surveys* for detailed information about survey development, content, and administration.)
3. Optional practice or training test content (Refer to subsection *5.2.2 Practice and Training Tests* for additional information.)

[1] The total population of students with the most significant cognitive disabilities in the California kindergarten through grade twelve public school system is approximately 38,000 (1 percent of the total student enrollment, which is provided in the CDE DataQuest website, for the 2015–2016 school year). Data was retrieved from https://dq.cde.ca.gov/dataquest/.

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 relating to the Science Connectors.

For the 2019–2020 CAASPP administration, embedded PTs were tested for the CAA for Science: four embedded PTs for grade five, four embedded PTs for grade eight, and four embedded PTs for high school (i.e., grade ten, eleven, or twelve). Each embedded PT included information for the test examiner, describing the hands-on activity and how to administer the embedded PT items. The embedded PT item types included selected-response, match, and grid items; these are described in subsection *3.1.3 Embedded Performance Task and Item Format*.

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

1.3.3.2. Survey on the Test Administration

During the 2019–2020 administration year, test examiners were asked to respond to a survey about their students. After each embedded PT was administered to the student, test examiners were presented with two surveys, with the instruction to only administer one of the surveys on the basis of whether or not their student had been responsive during the testing session. The purpose of the survey was to collect basic information about students' experiences with the assessment process.

Refer to chapter 8 for additional information about the student survey design.

1.3.3.3. Practice and Training Tests

Practice tests for each individual grade for all tested grade levels in all content domains were provided to LEAs to prepare students and LEA staff for the CAAs. A training embedded PT was also available in each tested grade level. Students, teachers, and the public may access them using a web browser.

These tests simulate the experience of the CAA online assessments and allow students and test examiners to become familiar with the user interface, item formats and functionality, available accessibility resources, and components of the TDS, as well as with the process of starting and completing a testing session. Unlike the summative CAA for Science, the practice and training tests do not assess standards, gauge student success on the operational assessment, or produce scores.

1.4. Intended Population

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

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

1.5. Intended Use and Purpose of the Test Scores

The results of tests within the CAASPP System, including the CAA for Science, are used for two primary purposes as described in *EC* sections 60602.5(a) and (a)(4). (Excerpted from the *EC* Section 60602 web page at https://leginfo.legislature.ca.gov/faces/codes_displayText.xhtml?lawCode=EDC&division=4.&title=2.&part=33.&chapter=5.&article=1 [outside source].)

> "60602.5(a) It is the intent of the Legislature in enacting this chapter to provide a system of assessments of pupils that has the primary purposes of assisting teachers, administrators, and pupils and their parents; improving teaching and learning; and promoting high-quality teaching and learning using a variety of assessment approaches and item types. The assessments, where applicable and valid, will produce scores that can be aggregated and disaggregated for the purpose of holding schools and local educational agencies accountable for the achievement of all their pupils in learning the California academic content standards."
>
> "60602.5(a)(4) Provide information to pupils, parents and guardians, teachers, schools, and local educational agencies on a timely basis so that the information can be used to further the development of the pupil and to improve the educational program."

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

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

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

> "60602.5(c) It is the intent of the Legislature that parents, classroom teachers, other educators, pupil representatives, institutions of higher education, business community members, and the public be involved, in an active and ongoing basis, in the design and implementation of the statewide pupil assessment system and the development of assessment instruments."
>
> "60602.5(d) It is the intent of the Legislature, insofar as is practically feasible and following the completion of annual testing, that the content, test structure, and test items in the assessments that are part of the statewide pupil assessment system become open and transparent to teachers, parents, and pupils, to assist stakeholders in working together to demonstrate improvement in pupil academic achievement. A planned change in annual test content, format, or design should be made available to educators and the public well before the beginning of the school year in which the change will be implemented."

However, no test scores were reported for the 2019–2020 administration because testing was suspended on March 18, 2020. More information on the suspension of testing is provided in the next section.

1.6. Testing Window and the Impact of the Novel Coronavirus Disease 2019 Pandemic

For the 2019–2020 CAASPP administration, the CAA for Science embedded PTs were available for administration on or after September 10, 2019, and were scheduled to be available for administration through the last day of instruction at the LEA or July 15, 2020, whichever came first (5 *CCR,* Section 855[a][2]). However, most of the schools in California halted in-person instruction after March 13, 2020. Then, on March 18, 2020, Governor Gavin Newsom signed an order suspending the CAASPP for all students in California (Office of Governor Gavin Newsom, 2020).

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

1.7. Significant Developments for the CAA for Science 2019–2020 Administration

The CAA for Science 2019–2020 administration had a number of significant developments. As the test moved to its first operational year, the number of embedded PTs that students took increased from three to four. The fourth embedded PT was comprised entirely of field test content so that students could have one operational embedded PT in each domain while still field testing items for future administrations. Additionally, the test itself used a functionality called the *Mark as No Response* option.

1.7.1. *Mark as No Response* Option

The *Mark as No Response* option is available on every item of the test within the TDS. If a student does not orient to an item, then the test examiner will select the *Mark as No Response* option, indicating that the student was presented with the item in question but did not respond. Previously, a test examiner would have just skipped ahead in the test, but this option allows for greater detail in the data gathered from the test; now it can be determined whether a student was presented an item.

1.7.2. Updated Accessibility Resources

The following changes were made to the list of CAA for Science accessibility resources:

- Streamline was reassigned as an embedded designated support.
- "Medical supports" replaced the name for the "medical device" as a non-embedded designated support for all assessments.

1.8. Impact of the Novel Coronavirus Disease 2019 Pandemic on Psychometric Analyses and Reporting

COVID-19 significantly impacted the 2019–2020 administration of the CAA for Science, as the testing window's suspension on March 18, 2020, occurred before most students were scheduled to test for this administration.

The number of test-takers was drastically lower than what was expected; only 367 grade five students, 435 grade eight students, and 512 high school students completed the CAA for Science even though over 4,000 students were registered to take each CAA for Science test. Due to the small numbers of students completing the CAA for Science ETS was unable to conduct the full data analyses—no classical item analyses, item response theory analyses, or reliability analyses were conducted. Additionally, without this data, standard setting could not be conducted as planned. The data gleaned from the students who did test is used in this report, but the sample is not representative of the population as a whole. Table 2.1 provides the number of students who were registered to take the CAA for Science and the percentage of students who completed the assessment.

To ensure the accuracy of all student scores, two members of the ETS Psychometric Analysis & Research team conducted independent, parallel checks of the percent correct and preliminary indicators.

1.9. Groups and Organizations Involved with the Assessment

1.9.1. State Board of Education

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

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

1.9.2. California Department of Education

The CDE oversees California's public school system, which is responsible for the education of more than 6,160,000 children and young adults in more than 10,500 schools.[2] California aims to provide a world-class education for all students, from early childhood to adulthood.

[2] Retrieved from the CDE Fingertip Facts on Education in California – *CalEdFacts* web page at https://www.cde.ca.gov/ds/ad/ceffingertipfacts.asp

The CDE serves the state by innovating and collaborating with educators, school staff, parents/guardians, and community partners which together, as a team, prepares students to live, work, and thrive in a highly connected world.

Within the CDE, it is the Instruction & Measurement branch that oversees programs promoting improved student achievement. Programs include oversight of statewide assessments and the collection and reporting of educational data (CDE, 2021).

1.9.3. 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, experience, demographics, and geographic locations—were invited to participate in the entire CAA for Science assessment development process. This included defining the purpose and scope of the assessment, assessment design, item development, data review, and score reporting.

1.9.4. Contractors

1.9.4.1. Primary Testing Contractor—ETS

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

- Providing management of the program activities
- Supporting and training 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 *DFAs*
- Processing student test assignments
- Completing all psychometric procedures
- Producing and distributing score reports
- Developing a score reporting website that can be viewed by the public

1.9.4.2. Subcontractor—Cambium Assessment, Inc.

ETS also monitors and manages the work of Cambium Assessment Inc. (CAI; formerly American Institutes for Research), subcontractor to ETS for the CAASPP System of online assessments. Activities conducted by CAI include

- providing the CAI proprietary TDS, including the Student Testing Interface, Test Administrator Interface, secure browser, and training tests;

- hosting and providing support for its TDS and the Online Reporting System (ORS),[3] a component of the overall CAASPP Assessment Delivery System;
- scoring machine-scorable items; and
- providing high-level technology help desk support to LEAs for technology issues directly related to the TDS.

1.10. Systems Overview and Functionality

1.10.1. Test Operations Management System

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

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

TOMS receives student enrollment data, including LEA and school hierarchy data, from the California Longitudinal Pupil Achievement Data System (CALPADS) via a daily feed. CALPADS is "a longitudinal data system used to maintain individual-level data including student demographics, course data, discipline, assessments, staff assignments, and other data for state and federal reporting."[4] LEA staff involved in the administration of the CAA for Science—such as LEA CAASPP coordinators, CAASPP test site coordinators, and test examiners—are assigned varying levels of access to TOMS. For example, only an LEA CAASPP coordinator is given permission to set up the LEA's test administration window; a test examiner cannot download student reports. A description of user roles is explained more extensively in the *2019–2020 CAASPP Online Test Administration Manual* (CDE, 2020a).

1.10.2. Test Delivery System

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

- the Test Administrator Interface, the web browser–based application that allows test administrators and test examiners to activate student tests and monitor student testing;

[3] The ORS will be replaced with the California Educator Reporting System (CERS) starting in January 2021.

[4] From the CDE California Longitudinal Pupil Achievement Data System (CALPADS) web page at https://www.cde.ca.gov/ds/sp/cl/

- the Student Testing Interface, on which students take the CAA for Science using the secure browser and with assistance from the test examiner as needed; and
- the secure browser, the online application through which the Student Testing Interface may be accessed. The secure browser prevents students from accessing other applications during testing.

1.10.3. Practice and Training Tests

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

The purposes of the practice and training tests are to

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

Details on practice and training tests are presented in subsection *5.2.1 Practice and Training Tests*.

1.11. Overview of the Technical Report

This technical report addresses the characteristics of the CAAs for Science administered from September 2019 through March 2020 and contains nine additional chapters as follows:

- Chapter 2 presents an overview of processes involved in the CAA for Science, including descriptions of item development, test administration, and psychometric analyses.
- Chapter 3 discusses the detailed procedures of embedded PT development for the CAA for Science.
- Chapter 4 describes the process of test assembly for the CAA for Science.
- Chapter 5 describes the details of administering the embedded PTs for the CAA for Science, as well as the procedures followed by ETS to ensure test security.
- Chapter 6 summarizes the scoring approaches and type of scores that are reported for the CAA for Science.
- Chapter 7 summarizes the statistical procedures and results for 2018–2019, including classical item analyses, test completion rates and analyses, and differential item functioning analyses.
- Chapter 8 describes the development and administration of the survey questionnaires for test examiners and the results of analyses conducted on their responses.

- Chapter 9 discusses the various procedures used to ensure the quality of the CAA for Science.
- Chapter 10 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. https://govt.westlaw.com/calregs/Document/I7907D7786C424BEFAC22247312385DFC?viewType=FullText&originationContext=documenttoc&transitionType=CategoryPageItem&contextData=(sc.Default)

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx

California Department of Education. (2020a). *CAASPP online test administration manual, 2019–2020 test administration.* Sacramento, CA: California Department of Education. https://bit.ly/3eXYBXc

California Department of Education. (2020b). *California Alternate Assessment for Science*. Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/caascience.asp

California Department of Education. (2020c). *SBE responsibilities.* https://www.cde.ca.gov/be/ms/po/sberesponsibilities.asp

California Department of Education. (2021). *Organization.* https://www.cde.ca.gov/re/di/or/

Gong, B., & Marion, S. (2006). Dealing with flexibility in assessments for students with significant cognitive disabilities. *Synthesis Report, 60*. https://education.umn.edu/NCEO/OnlinePubs/Synthesis60.html

Office of Governor Gavin Newsom. (2020). *Governor Newsom issues executive order to suspend standardized testing for students in response to COVID-19 outbreak* [Press release]. https://www.gov.ca.gov/2020/03/18/governor-newsom-issues-executive-order-to-suspend-standardized-testing-for-students-in-response-to-covid-19-outbreak/

Chapter 2: An Overview of the CAA for Science Processes

This chapter provides an overview of processes implemented by ETS during the full testing cycle for the 2019–2020 California Alternate Assessment (CAA) for Science administration, including descriptions of item development, test administration, and accessibility resources.

2.1. Embedded Performance Task and Item Development and Review

As part of the adaptation and alignment process, ETS developed all embedded performance tasks (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 Performance Task Development

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

2.1.2. Embedded Performance Task Development for Grades Five and Eight and High School

ETS developed each embedded PT with two sets of items, each set assessing a particular Science Connector. The concepts or topics that serve as the context for each item were reviewed to ensure that the content and presentation were accessible to, and developmentally appropriate for, students with the most significant cognitive disabilities.

A full review of the process to develop embedded PTs, including the number of items and the type of items, can be found in chapter 3.

2.1.2.1. Item Format

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

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

Detailed information on item format is included in subsection *3.1.4 Embedded Performance Task and Item Format* in *Chapter 3: Embedded Performance Task and Item Development and Review*.

SR and TEIs have either one or two points and are machine-scored.

2.1.2.2. Item Specifications

The CAA item specifications provide descriptions of item characteristics that are intended to measure each content standard consistently. They were developed based on the CA NGSS guidelines and clarifications from the Science Connectors; the focal knowledge, skills, and abilities (FKSAs); and essential understandings (EUs). During item development, item developers were provided with CAA item specifications and a CAA style guide that contained detailed information about the consistency in item development and item review processes. Refer to subsection *3.1.2 Specifications for the Embedded Performance Tasks and Items* in chapter 3 for detailed information about item specifications.

2.1.2.3. Item Banking

Typically, after each CAA for Science administration, item analyses are implemented and the results are reviewed by ETS psychometric and assessment development staff, who provide recommendations to the CDE on whether the items should be included or excluded in the pool of items for future administrations. Content experts from ETS and the CDE, as well as selected California educators, usually review the associated item statistics and evaluate the performance of items during the annual data review meeting. They also review the flagged items—those whose statistics fall beyond expected ranges—and work to provide plausible explanations for these particular items based on their knowledge of the student population.

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

2.1.3. Universal Design Principles

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

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

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

2.2. Test Assembly

The 2019–2020 operational assessment was assembled in accordance with the CAA for Science blueprint, which was approved by the SBE in January 2018 (CDE, 2018). The CAA for Science is a linear form comprised of three embedded PTs, each comprised of two Connector sets that assess Science Connectors from one of the three science domains of Life Sciences, Physical Sciences, and Earth and Space Sciences.

The assembly began with selection of approved anchor items from the item bank. For each embedded PT, a Connector set of five anchor items was paired with a Connector set of five operational field test items. After the initial assembly, test developers reviewed the assembled forms using comprehensive checklists to evaluate blueprint alignment, item content, clueing and content overlap, and overall balance of content with regard to gender and ethnicity representation, variety of item types, and so forth.

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

2.2.1. Test Design

The CAA for Science is based on a linear design comprised of three operational embedded PTs, each comprised of two Connector sets that assess standards from one of the three science domains. The Connector sets also incorporate contexts aligned to the Engineering, Technology, and Applications of Science domain. There is an additional, fourth embedded PT of field test items that do not count toward the student's total raw score.

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

The four embedded PTs—three operational embedded PTs and one field test embedded PT—were intended to be administered throughout the school year, shortly after students received instruction in the Science Connectors assessed by the embedded PT. Thus, the embedded PTs could be administered in any order throughout the instructional year.

The operational embedded PTs were available for administration from September 10, 2019, through March 18, 2020, the date when testing was suspended, per the governor's emergency order, because of the novel coronavirus disease 2019 (COVID-19) pandemic (Office of Governor Gavin Newsom, 2020).

2.2.2. Test Blueprints

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

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

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

2.2.3. Test Length

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

Refer to section *4.1 Overview* in *Chapter 4: Test Assembly* for more details on test form assembly.

2.3. Test Administration

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

2.3.1. Test Security and Confidentiality

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

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

- Standardization of test security
- Security of electronic files using a firewall
- Transfer of scores via secure data exchange
- Data management

- Statistical analysis
- Student confidentiality

2.3.2. Procedures to Maintain Standardization

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

2.3.2.1. Test Administration

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

All staff at LEAs that are involved in the CAASPP administration, including CAA for Science administration, are provided with directions about their responsibilities. Their roles include LEA CAASPP coordinators, CAASPP test site coordinators, and CAA test examiners. The responsibilities of each of the staff members specifically for the CAAs are described in the *2019–2020 CAASPP Online Test Administration Manual* (CDE, 2020c).

2.3.2.2. Test Directions

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

- ***CAA for Science DFAs—***The secure, grade-level *DFAs* are manuals that provide the script to be followed exactly by test examiners during a testing session. The *DFAs* for the CAA for Science contain embedded PT–specific instructions. Each version of each grade-level CAA for Science has a *DFA* for each embedded PT for that grade level. (Refer to *5.4.4.1 Directions for Administration* in chapter 5 for more information.)
- ***CAASPP Online Test Administration Manual—***This is a manual that provides test administration procedures and guidelines for LEA CAASPP coordinators, CAASPP test site coordinators, test examiners, and test administrators (CDE, 2020c). (Refer to *5.4.4.2 CAASPP Online Test Administration Manual* in chapter 5 for more information.)
- ***CAASPP and English Language Proficiency Assessments for California (ELPAC) Test Operations Management System (TOMS) User Guide—***This is a manual that provides instructions for TOMS allowing LEA staff, including LEA CAASPP coordinators and CAASPP test site coordinators, to perform a number of tasks including setting up test administrations, adding and managing users, assigning tests, configuring online student test settings, and accessing the secure *DFAs* (CDE, 2020b). (Refer to *5.4.4.3 CAASPP and English Language Proficiency Assessments for California (ELPAC) TOMS User Guide* in chapter 5 for more information.)

2.4. Test-Taking Rates

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

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

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

All students who were identified to take the CAAs were required to test.

Table 2.1 presents the number of test takers assigned to take the CAA for Science and the number of students who started the CAA for Science. Table 2.1 also presents the number of students whose test expired, whose test was force-completed, or who submitted all four embedded PTs. Students with an expired test started one or more embedded PTs but did not complete the started embedded PT(s); these embedded PTs that were not submitted by a student were submitted for processing by the system. Students who had their tests force-completed had unused embedded PTs that required additional, manual steps to submit for processing after the end of the statewide testing window.

The majority of students assigned to take the CAA for Science had not started any of the embedded PTs when testing was suspended on March 18, 2020 per the governor's emergency order, because of the COVID-19 pandemic (Office of Governor Gavin Newsom, 2020). For grades five and eight, approximately half of the students who started the CAA for Science completed all four embedded PTs. Across the high school grades, more than half of the students who started the test completed all four embedded PTs.

Table 2.1 CAA for Science Test-Taking Rates—Registered Students

Group	Grade 5	Grade 8	HS Grade 10	HS Grade 11	HS Grade 12	HS All Grades
Number Assigned	4,249	4,622	392	2,832	1,546	4,770
Number Started	737	804	136	347	246	729
Percent Started	17%	17%	35%	12%	16%	15%
Number of Force-Completed	365	364	28	132	54	214
Percent Force-Completed	9%	8%	7%	5%	3%	4%
Number of Expired	5	5	1	0	2	3
Percent Expired	0%	0%	0%	0%	0%	0%
Number of Submitted	367	435	107	215	190	512
Percent Submitted	9%	9%	27%	8%	12%	11%

Note: The percentages of students with force-completed, expired, or submitted tests may not sum to the percentage of students who started the tests because of rounding.

Table 2.2 presents the percentage of students in each grade level or grade band that completed one, two, three, four, or none of the embedded PTs. The majority of registered students did not complete any of the four the CAA for Science embedded PTs by the time testing was suspended on March 18, 2020.

Table 2.2 Percentage of Students in Each Grade Level or Grade Band Completing Embedded PTs

Grade Level or Grade Band	No PTs Completed	1 PT Completed	2 PTs Completed	3 PTs Completed	4 PTs Completed	Number of Students Assigned
Grade 5	83%	3%	4%	2%	9%	4,249
Grade 8	83%	4%	3%	2%	9%	4,622
High school	85%	1%	2%	1%	11%	4,770
Grade 10	65%	1%	5%	2%	27%	392
Grade 11	88%	1%	2%	1%	8%	2,832
Grade 12	84%	1%	1%	1%	12%	1,546

2.5. Fairness and Accessibility

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

2.5.1. Universal Tools, Designated Supports, and Accommodations

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

The CDE maintains a list of the universal tools, designated supports, and accommodations that are permitted for use in CAASPP online assessments in its web document *Matrix One: California Assessment of Student Performance and Progress Accessibility Resources* (CDE, 2019b).[5]

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

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

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

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

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

[5] This technical report is based on the version of Matrix One that was available during the 2019–2020 CAASPP administration. Note that Matrix One has since been combined with the ELPAC Matrix Four to form a single accessibility resources matrix, the California Assessment Accessibility Resources Matrix (CDE, 2020d).

Table 2.A.1 through Table 2.A.6 in appendix 2.A present the number and percentages of students using designated supports, accommodations, or unlisted resources for the 2019–2020 CAA for Science administration. The use of universal tools is not tracked because they are available to all students in the TDS.

These tables are provided for each of the three operational embedded PTs for each science domain. All students who completed the embedded PT and used at least one of the designated supports or accommodations are included in these tables. Because CAA for Science testing was suspended on March 18, 2020, not all students were given the opportunity to complete their tests. Thus, students who completed an embedded PT but who may not have a valid test score because testing was suspended are included in these tables.

Refer to section *1.9 Systems Overview and Functionality* in *Chapter 1: Introduction* for more details regarding these systems.

2.5.1.1. Resources for Selection of Accessibility Resources

The full list of the universal tools, designated supports, and accommodations that are used in CAASPP online assessments, including the CAA for Science, are documented in Matrix One (CDE, 2019b). Most embedded and non-embedded universal tools, designated supports, and accommodations listed in parts 1, 2, and 3 of Matrix One are available for the CAA for Science through the online testing interface or, in the case of non-embedded resources, from the school or LEA. Part 4 of Matrix One includes instructional supports and resources available for a student taking an alternate assessment.

School-level personnel and IEP teams used Matrix One when deciding how best to support the student's test-taking experience. IEP teams may consider what other designated supports and accommodations, other than universal resources already available in the TDS, may be appropriate for the student. On the rare occasion when a student has both an IEP and a Section 504 plan, the Section 504 plan also should be referenced for accessibility resources.

In addition to assigning accessibility resources individually and via file upload in TOMS, LEAs had the option of using the ISAAP Tool to assign resources to students. The ISAAP Tool was used by LEAs in conjunction with the *Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Guidelines* (Smarter Balanced, 2020)[6] *and the 2019–2020 CAASPP and ELPAC Accessibility Guide for Online Testing* (CDE, 2020a), as well as with state regulations and policies (such as Matrix One) related to assessment accessibility.

2.5.1.2. Delivery of Accessibility Resources

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

[6] This technical report is based on the version of the *Usability, Accessibility, and Accommodations Guidelines* that was available during the 2019–2020 CAASPP administration.

2.5.1.3. Unlisted Resources

An unlisted resource is an instructional support, identified in the student's IEP, that a student regularly uses in daily instruction, assessment, or both, and has not been previously identified as a universal tool, designated support, or accommodation. Matrix One included an inventory of unlisted resources that were already identified and were preapproved (CDE, 2019b). During the 2019–2020 CAASPP administration, an LEA CAASPP coordinator or a CAASPP test site coordinator would use TOMS to submit a request for use of an unlisted resource. A request for an unlisted resource that was not preidentified was sent to the CDE for approval.

If a student used an unlisted resource that changed the construct of the assessment, the student's score was invalidated.

2.5.2. 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 section *5.5 Accessibility Features for the 2019–2020 Administration*.

2.5.3. Description of Differential Item Functioning Analyses

DIF analyses are typically conducted to detect differences in student performance by identifying items on which one group of students performs significantly better than another group (e.g., male vs. female or White vs. African-American) after matching students on ability. If an item performed differentially across student groups, even when students were matched on ability, the item may be measuring something other than the intended construct. Therefore, it is important to identify items flagged for DIF. Content experts and bias and sensitivity experts review these DIF-flagged items to determine the sources and meanings of performance differences. Refer to section *7.5. Differential Item Functioning Analyses* for DIF analyses conducted.

2.6. Scores

Student responses to each embedded PT were machine-scored.

2.6.1. Score Reporting

There were no individual student scores reported for the 2019–2020 CAA for Science administration because of the suspension of testing on March 18, 2020. ETS prepared a data file of students' percent-correct scores and the associated preliminary indicator category for LEAs.

2.7. Overview of Psychometric Analyses

There were a number of psychometric analyses planned for the CAA for Science data, including classical item analyses, DIF, item response theory calibrations, reliability, and response time analyses. These analyses are described fully in *Chapter 7: Psychometric Analyses*. Because of the COVID-19 pandemic and the suspension of testing on March 18, 2020, the planned psychometric analyses were not conducted for the 2019–2020 administration.

2.7.1. Description of the Classical Item Analyses

The psychometric analyses for the CAA for Science typically consist of classical item analyses and DIF analyses to evaluate the performance of the embedded PT items. The classical item analyses include the computation of item difficulty indices, the item-total correlation indices, the omit rate of each embedded PT item, and the proportion of test takers obtaining each score point for the polytomous items. Flagging rules based on these statistics identify items not performing as expected. Descriptions of the typical psychometric analyses are provided in section *7.2 Classical Item Analyses*.

2.7.2. Description of Item Response Theory Analyses

Typically, a concurrent calibration is conducted typically to estimate parameters for all items. As a result of the concurrent calibration, the item parameter estimates are placed on a common scale for test items from the same grade-level test. The concurrent calibration requires either “common items” or “random equivalent groups.”

The CAA for Science versions are assembled with common items between the versions, which support the efficiency and accuracy of the concurrent calibrations. The one-parameter logistic item response theory model (Hambleton and Rogers, 1991) and the partial credit model (Masters, 1982) are used for item calibration of the CAA for Science with flexMIRT® (Cai, 2016) version 3.5 software.

Detailed procedures for the concurrent calibrations are included in subsection *7.6.2. Item Calibration*.

References

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

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx

California Department of Education. (2019a). *Alternate assessment IEP team guidance.* https://www.cde.ca.gov/ta/tg/ca/caaiepteamrev.asp

California Department of Education. (2019b). *Matrix one: California Assessment of Student Performance and Progress accessibility resources.* Sacramento, CA: California Department of Education. https://bit.ly/3h0EkD2

California Department of Education. (2020a) *CAASPP and ELPAC accessibility guide for online testing.* Sacramento, CA: California Department of Education. https://bit.ly/3usOzny

California Department of Education. (2020b). *CAASPP and ELPAC Test Operations Management System user guide*. Sacramento, CA: California Department of Education. https://bit.ly/2QKKJaV

California Department of Education. (2020c). *CAASPP online test administration manual, 2019–2020 test administration.* Sacramento, CA: California Department of Education. https://bit.ly/3eXYBXc

California Department of Education. (2020d). *California assessment accessibility resources matrix*. Sacramento, CA: California Department of Education. https://bit.ly/3vHmKbm

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

Office of Governor Gavin Newsom. (2020). *Governor Newsom issues executive order to suspend standardized testing for students in response to COVID-19 outbreak* [Press release]. https://www.gov.ca.gov/2020/03/18/governor-newsom-issues-executive-order-to-suspend-standardized-testing-for-students-in-response-to-covid-19-outbreak/

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

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

Smarter Balanced Assessment Consortium. (2020). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations guidelines.* Los Angeles, CA: Smarter Balanced Assessment Consortium. https://portal.smarterbalanced.org/library/en/usability-accessibility-and-accommodations-guidelines.pdf

Appendix 2.A: Accessibility

Table 2.A.1 Assignment of Designated Supports and Accommodations—Grades Five and Eight for Earth and Space Sciences

Accessibility Resource	Grade 5: N	Grade 5: % of Total Tested	Grade 8: N	Grade 8: % of Total Tested
Non-Embedded Accommodation—Abacus	7	1%	6	1%
Non-Embedded Accommodation—Additional Instructional Supports for Alternate Assessments	38	5%	55	7%
Non-Embedded Accommodation—Alternate Response Options	34	5%	66	8%
Non-Embedded Accommodation—Print on Demand	5	1%	17	2%
Non-Embedded Accommodation—Unlisted Resources	0	0%	0	0%
Non-Embedded Accommodation—Word Prediction	4	1%	22	3%
Embedded Designated Support—Color Contrast	3	0%	5	1%
Embedded Designated Support—Masking	14	2%	35	4%
Embedded Designated Support—Mouse Pointer	8	1%	3	0%
Embedded Designated Support—Permissive Mode	6	1%	16	2%
Embedded Designated Support—Print Size	13	2%	16	2%
Embedded Designated Support—Streamline	12	2%	17	2%
Embedded Designated Support—Turn Off Any Universal Tools	0	0	0	0

Table 2.A.1 *(continuation)*

Accessibility Resource	Grade 5: N	Grade 5: % of Total Tested	Grade 8: N	Grade 8: % of Total Tested
Non-Embedded Designated Support—Amplification	3	0%	2	0%
Non-Embedded Designated Support—Color Contrast	1	0%	6	1%
Non-Embedded Designated Support—Color Overlay	1	0%	6	1%
Non-Embedded Designated Support—Magnification	8	1%	15	2%
Non-Embedded Designated Support—Medical Supports	1	0%	0	0%
Non-Embedded Designated Support—Multiplication Table	24	3%	42	5%
Non-Embedded Designated Support—Noise Buffers	33	4%	52	6%
Non-Embedded Designated Support—Read Aloud Items	113	15%	186	23%
Non-Embedded Designated Support—Scribe Items	39	5%	88	11%
Non-Embedded Designated Support—Separate Setting	131	18%	186	23%
Non-Embedded Designated Support—100s Number Table	28	4%	41	5%
Total Students Tested	**737**	**N/A**	**804**	**N/A**

Note: Some students are eligible for multiple accessibility resources. As a result, the number of students tested in each grade level may not equal the sum of the number of students eligible per accessibility resource across all accessibility resources.

Table 2.A.2 Assignment of Designated Supports and Accommodations—Grades Five and Eight for Physical Sciences

Accessibility Resource	Grade 5: N	Grade 5: % of Total Tested	Grade 8: N	Grade 8: % of Total Tested
Non-Embedded Accommodation—Abacus	3	0%	6	1%
Non-Embedded Accommodation—Additional Instructional Supports for Alternate Assessments	41	6%	43	5%
Non-Embedded Accommodation—Alternate Response Options	25	3%	60	7%
Non-Embedded Accommodation—Print on Demand	2	0%	8	1%
Non-Embedded Accommodation—Unlisted Resources	0	0%	0	0%
Non-Embedded Accommodation—Word Prediction	2	0%	15	2%
Embedded Designated Support—Color Contrast	2	0%	2	0%
Embedded Designated Support—Masking	17	2%	26	3%
Embedded Designated Support—Mouse Pointer	4	1%	3	0%
Embedded Designated Support—Permissive Mode	3	0%	10	1%
Embedded Designated Support—Print Size	7	1%	13	2%
Embedded Designated Support—Streamline	9	1%	13	2%
Embedded Designated Support—Turn Off Any Universal Tools	0	0	0	0

Table 2.A.2 *(continuation)*

Accessibility Resource	Grade 5: N	Grade 5: % of Total Tested	Grade 8: N	Grade 8: % of Total Tested
Non-Embedded Designated Support—Amplification	4	1%	3	0%
Non-Embedded Designated Support—Color Contrast	1	0%	4	1%
Non-Embedded Designated Support—Color Overlay	1	0%	4	1%
Non-Embedded Designated Support—Magnification	4	1%	15	2%
Non-Embedded Designated Support—Medical Supports	1	0%	0	0%
Non-Embedded Designated Support—Multiplication Table	13	2%	27	3%
Non-Embedded Designated Support—Noise Buffers	32	4%	39	5%
Non-Embedded Designated Support—Read Aloud Items	86	12%	154	19%
Non-Embedded Designated Support—Scribe Items	27	4%	75	9%
Non-Embedded Designated Support—Separate Setting	93	13%	154	19%
Non-Embedded Designated Support—100s Number Table	15	2%	31	4%
Total Students Tested	737	N/A	804	N/A

Note: Some students are eligible for multiple accessibility resources. As a result, the number of students tested in each grade level may not equal the sum of the number of students eligible per accessibility resource across all accessibility resources.

Table 2.A.3 Assignment of Designated Supports and Accommodations—Grades Five and Eight for Life Sciences

Accessibility Resource	Grade 5: N	Grade 5: % of Total Tested	Grade 8: N	Grade 8: % of Total Tested
Non-Embedded Accommodation—Abacus	7	1%	6	1%
Non-Embedded Accommodation—Additional Instructional Supports for Alternate Assessments	49	7%	43	5%
Non-Embedded Accommodation—Alternate Response Options	34	5%	64	8%
Non-Embedded Accommodation—Print on Demand	0	0%	12	1%
Non-Embedded Accommodation—Unlisted Resources	0	0%	0	0%
Non-Embedded Accommodation—Word Prediction	2	0%	22	3%
Embedded Designated Support—Color Contrast	2	0%	4	1%
Embedded Designated Support—Masking	21	3%	34	4%
Embedded Designated Support—Mouse Pointer	5	1%	4	1%
Embedded Designated Support—Permissive Mode	4	1%	10	1%
Embedded Designated Support—Print Size	11	1%	16	2%
Embedded Designated Support—Streamline	13	2%	18	2%
Embedded Designated Support—Turn Off Any Universal Tools	1	0%	1	0%

Table 2.A.3 *(continuation)*

Accessibility Resource	Grade 5: N	Grade 5: % of Total Tested	Grade 8: N	Grade 8: % of Total Tested
Non-Embedded Designated Support—Amplification	4	1%	3	0%
Non-Embedded Designated Support—Color Contrast	1	0%	6	1%
Non-Embedded Designated Support—Color Overlay	2	0%	6	1%
Non-Embedded Designated Support—Magnification	5	1%	16	2%
Non-Embedded Designated Support—Medical Supports	1	0%	0	0%
Non-Embedded Designated Support—Multiplication Table	24	3%	39	5%
Non-Embedded Designated Support—Noise Buffers	41	6%	51	6%
Non-Embedded Designated Support—Read Aloud Items	117	16%	174	22%
Non-Embedded Designated Support—Scribe Items	39	5%	82	10%
Non-Embedded Designated Support—Separate Setting	132	18%	173	22%
Non-Embedded Designated Support—100s Number Table	28	4%	40	5%
Total Students Tested	**737**	**N/A**	**804**	**N/A**

Note: Some students are eligible for multiple accessibility resources. As a result, the number of students tested in each grade level may not equal the sum of the number of students eligible per accessibility resource across all accessibility resources.

Table 2.A.4 Assignment of Designated Supports and Accommodations—High School for Earth and Space Sciences

Accessibility Resource	Grade 10: N	Grade 10: % of Total Tested	Grade 11: N	Grade 11: % of Total Tested	Grade 12: N	Grade 12: % of Total Tested	High School: N	High School: % of Total
Non-Embedded Accommodation—Abacus	0	0%	2	1%	1	0%	3	0%
Non-Embedded Accommodation—Additional Instructional Supports for Alternate Assessments	7	5%	22	6%	6	2%	35	5%
Non-Embedded Accommodation—Alternate Response Options	11	8%	9	3%	7	3%	27	4%
Non-Embedded Accommodation—Print on Demand	0	0%	0	0%	0	0%	0	0%
Non-Embedded Accommodation—Unlisted Resources	0	0%	0	0%	0	0%	0	0%
Non-Embedded Accommodation—Word Prediction	1	1%	7	2%	1	0%	9	1%
Embedded Designated Support—Color Contrast	0	0%	0	0%	0	0%	0	0%
Embedded Designated Support—Masking	8	6%	1	0%	4	2%	13	2%
Embedded Designated Support—Mouse Pointer	1	1%	2	1%	2	1%	5	1%
Embedded Designated Support—Permissive Mode	2	1%	3	1%	1	0%	6	1%
Embedded Designated Support—Print Size	2	1%	2	1%	2	1%	6	1%
Embedded Designated Support—Streamline	3	2%	1	0%	4	2%	8	1%
Embedded Designated Support—Turn Off Any Universal Tools	0	0	0	0	0	0	0	0

Table 2.A.4 *(continuation)*

Accessibility Resource	Grade 10: N	Grade 10: % of Total Tested	Grade 11: N	Grade 11: % of Total Tested	Grade 12: N	Grade 12: % of Total Tested	High School: N	High School: % of Total
Non-Embedded Designated Support—Amplification	0	0%	1	0%	1	0%	2	0%
Non-Embedded Designated Support—Color Contrast	0	0%	0	0%	2	1%	2	0%
Non-Embedded Designated Support—Color Overlay	0	0%	0	0%	1	0%	1	0%
Non-Embedded Designated Support—Magnification	1	1%	4	1%	2	1%	7	1%
Non-Embedded Designated Support—Medical Supports	0	0%	0	0%	1	0%	1	0%
Non-Embedded Designated Support—Multiplication Table	1	1%	17	5%	4	2%	22	3%
Non-Embedded Designated Support—Noise Buffers	8	6%	13	4%	12	5%	33	5%
Non-Embedded Designated Support—Read Aloud Items	33	24%	34	10%	52	21%	119	16%
Non-Embedded Designated Support—Scribe Items	7	5%	10	3%	14	6%	31	4%
Non-Embedded Designated Support—Separate Setting	31	23%	44	13%	46	19%	121	17%
Non-Embedded Designated Support—100s Number Table	5	4%	16	5%	3	1%	24	3%
Total Students Tested	**136**	**N/A**	**347**	**N/A**	**246**	**N/A**	**729**	**N/A**

Note: Some students are eligible for multiple accessibility resources. As a result, the number of students tested in each grade level may not equal the sum of the number of students eligible per accessibility resource across all accessibility resources.

Table 2.A.5 Assignment of Designated Supports and Accommodations—High School for Physical Sciences

Accessibility Resource	Grade 10: N	Grade 10: % of Total Tested	Grade 11: N	Grade 11: % of Total Tested	Grade 12: N	Grade 12: % of Total Tested	High School: N	High School: % of Total
Non-Embedded Accommodation—Abacus	0	0%	1	0%	1	0%	2	0%
Non-Embedded Accommodation—Additional Instructional Supports for Alternate Assessments	5	4%	11	3%	7	3%	23	3%
Non-Embedded Accommodation—Alternate Response Options	8	6%	5	1%	8	3%	21	3%
Non-Embedded Accommodation—Print on Demand	0	0%	0	0%	0	0%	0	0%
Non-Embedded Accommodation—Unlisted Resources	0	0%	0	0%	0	0%	0	0%
Non-Embedded Accommodation—Word Prediction	1	1%	1	0%	1	0%	3	0%
Embedded Designated Support—Color Contrast	0	0%	0	0%	0	0%	0	0%
Embedded Designated Support—Masking	7	5%	0	0%	4	2%	11	2%
Embedded Designated Support—Mouse Pointer	1	1%	2	1%	2	1%	5	1%
Embedded Designated Support—Permissive Mode	0	0%	1	0%	1	0%	2	0%
Embedded Designated Support—Print Size	1	1%	1	0%	2	1%	4	1%
Embedded Designated Support—Streamline	1	1%	0	0%	4	2%	5	1%
Embedded Designated Support—Turn Off Any Universal Tools	0	0	0	0	0	0	0	0

Table 2.A.5 *(continuation)*

Accessibility Resource	Grade 10: N	Grade 10: % of Total Tested	Grade 11: N	Grade 11: % of Total Tested	Grade 12: N	Grade 12: % of Total Tested	High School: N	High School: % of Total
Non-Embedded Designated Support—Amplification	0	0%	0	0%	1	0%	1	0%
Non-Embedded Designated Support—Color Contrast	0	0%	0	0%	2	1%	2	0%
Non-Embedded Designated Support—Color Overlay	0	0%	0	0%	1	0%	1	0%
Non-Embedded Designated Support—Magnification	0	0%	4	1%	2	1%	6	1%
Non-Embedded Designated Support—Medical Supports	0	0%	0	0%	1	0%	1	0%
Non-Embedded Designated Support—Multiplication Table	0	0%	5	1%	4	2%	9	1%
Non-Embedded Designated Support—Noise Buffers	8	6%	5	1%	12	5%	25	3%
Non-Embedded Designated Support—Read Aloud Items	30	22%	22	6%	50	20%	102	14%
Non-Embedded Designated Support—Scribe Items	7	5%	7	2%	10	4%	24	3%
Non-Embedded Designated Support—Separate Setting	26	19%	25	7%	49	20%	100	14%
Non-Embedded Designated Support—100s Number Table	4	3%	5	1%	3	1%	12	2%
Total Students Tested	**136**	**N/A**	**347**	**N/A**	**246**	**N/A**	**729**	**N/A**

Note: Some students are eligible for multiple accessibility resources. As a result, the number of students tested in each grade level may not equal the sum of the number of students eligible per accessibility resource across all accessibility resources.

Table 2.A.6 Assignment of Designated Supports and Accommodations—High School for Life Sciences

Accessibility Resource	Grade 10: N	Grade 10: % of Total Tested	Grade 11: N	Grade 11: % of Total Tested	Grade 12: N	Grade 12: % of Total Tested	High School: N	High School: % of Total
Non-Embedded Accommodation—Abacus	0	0%	2	1%	1	0%	3	0%
Non-Embedded Accommodation—Additional Instructional Supports for Alternate Assessments	7	5%	18	5%	7	3%	32	4%
Non-Embedded Accommodation—Alternate Response Options	11	8%	8	2%	8	3%	27	4%
Non-Embedded Accommodation—Print on Demand	0	0%	0	0%	0	0%	0	0%
Non-Embedded Accommodation—Unlisted Resources	0	0%	0	0%	0	0%	0	0%
Non-Embedded Accommodation—Word Prediction	1	1%	7	2%	1	0%	9	1%
Embedded Designated Support—Color Contrast	0	0%	0	0%	0	0%	0	0%
Embedded Designated Support—Masking	8	6%	1	0%	4	2%	13	2%
Embedded Designated Support—Mouse Pointer	1	1%	2	1%	2	1%	5	1%
Embedded Designated Support—Permissive Mode	2	1%	2	1%	1	0%	5	1%
Embedded Designated Support—Print Size	2	1%	1	0%	2	1%	5	1%
Embedded Designated Support—Streamline	3	2%	1	0%	4	2%	8	1%
Embedded Designated Support—Turn Off Any Universal Tools	0	0%	1	0%	0	0%	1	0%

Table 2.A.6 *(continuation)*

Accessibility Resource	Grade 10: N	Grade 10: % of Total Tested	Grade 11: N	Grade 11: % of Total Tested	Grade 12: N	Grade 12: % of Total Tested	High School: N	High School: % of Total
Non-Embedded Designated Support—Amplification	0	0%	1	0%	1	0%	2	0%
Non-Embedded Designated Support—Color Contrast	0	0%	0	0%	2	1%	2	0%
Non-Embedded Designated Support—Color Overlay	0	0%	0	0%	1	0%	1	0%
Non-Embedded Designated Support—Magnification	1	1%	4	1%	2	1%	7	1%
Non-Embedded Designated Support—Medical Supports	0	0%	0	0%	1	0%	1	0%
Non-Embedded Designated Support—Multiplication Table	0	0%	18	5%	4	2%	22	3%
Non-Embedded Designated Support—Noise Buffers	8	6%	9	3%	13	5%	30	4%
Non-Embedded Designated Support—Read Aloud Items	32	24%	35	10%	51	21%	118	16%
Non-Embedded Designated Support—Scribe Items	7	5%	11	3%	14	6%	32	4%
Non-Embedded Designated Support—Separate Setting	30	22%	38	11%	47	19%	115	16%
Non-Embedded Designated Support—100s Number Table	4	3%	18	5%	3	1%	25	3%
Total Students Tested	**136**	**N/A**	**347**	**N/A**	**246**	**N/A**	**729**	**N/A**

Note: Some students are eligible for multiple accessibility resources. As a result, the number of students tested in each grade level may not equal the sum of the number of students eligible per accessibility resource across all accessibility resources.

Chapter 3: Embedded Performance Task and Item Development and Review

This chapter provides an overview of the processes implemented by 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).

3.1. Embedded Performance Task and Item Development

3.1.1. Overview

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.

3.1.2. Specifications for the Embedded Performance Tasks and Items

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

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

3.1.3. Five-Year Plan

The CAA for Science blueprints require that all of the Science Connectors prioritized for assessment be assessed at least once during a five-year period (CDE, 2018). To support the planning for this rotation, ETS and the CDE collaborated to create a five-year plan. This plan contains a running record of the Science Connectors that have been assessed as well as a projection of the Science Connectors proposed for assessments in the coming years. The plan is reviewed and updated annually and is consulted during the planning for item development and forms construction.

3.1.4. Embedded Performance Task and Item Format

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

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

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

The CAA for Science included the following item formats:

- **Selected response (SR)—** A student was instructed to select one or more choices. Most CAA for Science items had two or three options; a few items had four options.
- **Match—**A student was instructed to place a picture on a specified part of a diagram or chart.
- **Grid—**A student was instructed to place a check mark in a specified section in a table of responses.

All SR, match, and grid items were scored by the test delivery system.

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

Table 3.1 Number of Items and Points for Each Embedded PT

Grade Level or Grade Band	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	Number of Items—Field Test PT	Number of Points—Field Test PT	Total Number of Operational Items	Maximum Number of Points
Grade 5	10	12	10	12	10	12	10	12	30	36
Grade 8	10	12	10	12	10	12	10	12	30	36
High School	10	12	10	12	10	12	10	12	30	36

Note: The field test embedded PT does not count towards a student's total raw score, the total number of operational items, or the maximum number of points.

3.1.5. Recruitment and Selection of Embedded Performance Task Item Writers

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

All item writers met the following minimum qualifications:

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

3.1.6. Embedded Performance Task Item Writer Training

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

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

3.2. ETS Item Review Process

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

3.2.1. Overview

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

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

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

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

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

3.2.2. ETS Content Review

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

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

Each embedded PT item was classified with the Science Connector, EU, and the FKSA it was intended to measure. Assessment specialists checked each item against its

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

3.2.3. ETS Editorial Review

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

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

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

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

3.3. California Educator Review

3.3.1. California Educators as Content Experts

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

These educators were responsible for reviewing all newly developed items for alignment to the CA NGSS and Science Connectors. Meeting participants also reviewed the items for accuracy of content, clarity of phrasing, and quality. In their examination of embedded PT items, participants can raise concerns about the appropriateness of the items as related to the grade, age, and cognitive level of the test taker. Additionally, items were evaluated for any potential bias or sensitivity concerns associated with disability, gender, race, ethnicity, religion, or socioeconomic status. ETS recorded educator feedback for each item and adjusted item content based on approval from the CDE.

3.3.2. Composition of Item Review Panels

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

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

Preferred qualifications included

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

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

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

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

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

Table 3.2 Number of Item Reviewers with Each Qualification

Qualification Type	Qualification	Total
N/A	**Total number of reviewers**	**9**
Occupation	Teacher or Program Specialist, Elementary School	3
Occupation	Teacher or Program Specialist, Middle School	4
Occupation	Teacher or Program Specialist, High School	2
Occupation	Other District Personnel	0
Highest Degree Earned	Bachelor's Degree	0
Highest Degree Earned	Master's Degree	9
Highest Degree Earned	Doctorate	1

Table 3.2 *(continuation)*

Qualification Type	Qualification	Total
K–12 Teaching Credential	Elementary Teaching (multiple subjects)	3
K–12 Teaching Credential	Secondary Teaching (single subject)	2
K–12 Teaching Credential	Special Education	5
K–12 Teaching Credential	Reading Specialist	0
K–12 Teaching Credential	English Learner (Crosscultural, Language and Academic Development; Bilingual, Crosscultural, Language and Academic Development)	0
K–12 Teaching Credential	Administrative	0
K–12 Teaching Credential	Other	0

Note: Numbers may not match the totals because item reviewers may have multiple occupations or teaching credentials or are currently working toward earning their highest degree.

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

3.3.3. Meetings for Review of CAA for Science Embedded Performance Tasks and Items

The 2019–2020 CAA for Science Item Review Meeting was held from to April 2 to April 5, 2019. ETS content-area assessment specialists facilitated CAA for Science item review meetings. Each meeting began with a brief training session on how to review and make recommendations for revising items. ETS provided training on the following topics:

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

The criteria for evaluating items included the following:

- Overall technical quality
- Alignment with the Science Connectors
- Alignment with the construct being assessed by the Science Connector
- Difficulty range
- Clarity
- Correctness of the answer
- Plausibility of the distractors
- Bias and sensitivity factors

Criteria also encompassed more global factors, including the quality of the alternative text to confirm that it describes an image in an age- and audience-appropriate manner within the

context of the question. Meeting participants also were trained on how to make recommendations for revising items.

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

- Does the item
  - have one and only one clearly correct answer (for single-select items)?
  - measure the content standard?
  - match the item specifications?
  - align with the construct being measured?
  - test worthwhile concepts or information?
- Is the stimulus, if any, for the item
  - required 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 Meeting

No data review meeting was held after the 2019–2020 CAA for Science administration because of the suspension of testing on March 18, 2020. Typically, after items are administered to students, ETS prepares the items and the associated statistics for review by the CDE and California educators.

In previous data review meetings for the CAA for Science, review materials included embedded PT items with their statistical data along with annotated comment sheets for use by reviewers. ETS conducted an introductory training to highlight any new issues and serve as a statistical refresher. Reviewers then made decisions about which items should be included in the item bank for future assembly. If an item was considered problematic and not to be included in the item bank, it could be revised, field-tested once again, and put through another round of item analysis. ETS psychometric and content staff were available to reviewers throughout this process.

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

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

Reference

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx

Chapter 4: Test Assembly

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

4.1. Overview

The operational California Alternate Assessment (CAA) for Science was administered as three embedded performance tasks (PTs) during the school year in each of grades five and eight and in high school. Each embedded PT within a grade assessed one of the three science domains, which are Life Sciences, Physical Sciences, and Earth and Space Sciences.

Each embedded PT assessed two Science Core Content Connectors (Science Connectors) from a domain. The embedded PTs contained an orienting activity and five questions aligned to each of the two Science Connectors. Thus, an embedded PT contained 10 questions assessing two Science Connectors, each with an orienting activity. Some of the orienting activities and questions incorporated simple activities designed to demonstrate a key concept associated with the assessed Science Connector.

4.2. Test Blueprints and Test Content Specifications

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

4.2.1. Test Blueprints

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

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

The 2019–2020 forms had 100 percent alignment with the test blueprints. Each of the three content domains were assessed by 10 items for a total of 12 points.

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

4.2.2. Test Content Specifications

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

Each content standard is assessed through the Science Connectors and related FKSAs and EUs under a three-level structure of item complexity.

4.3. Test Production Process

4.3.1. Selection of Items

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

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

4.3.2. Psychometric Criteria and Verification of Statistics

ETS test developers sent the proposed assessment to the ETS psychometrics team for approval. The proposed assessment was reviewed to ensure that all statistical guidelines were met for both individual items and the assessment as a whole. ETS psychometricians reviewed the item statistics, such as the *p*-value (item difficulty; refer to subsection *7.2.1 Classical Item Difficulty Indices (p-value and Average Item Score)* for more details on this statistic) and item-total correlation (item discrimination; refer to subsection *7.2.2 Item Discrimination (Item-Total Correlation)* for more details on this statistic) obtained from the field test administration, and used them to inform the item selection for the operational forms. At the form level, the distribution of *p*-values ranged from 0.35 to 0.92, and all the items selected had item-total correlations higher than 0.30.

The following psychometric criteria were applied in the form assembly:

- The *p*-value is between 0.2 and 0.95. A *p*-value less than 0.2 suggests that the item might be too difficult; a *p*-value greater than 0.95 suggests that the item might be too easy. Items that were too easy or too difficult were not used, as they provided little information on evaluating students' abilities.
- The item-total correlation is at least 0.2. Items selected had item-total correlations higher than 0.3.

- Items with C-DIF should not be used unless it is necessary for content coverage (refer to section *7.5 Differential Item Functioning Analyses* for more details on the differential item functioning [DIF]). All C-DIF items were reviewed by a DIF panel that included members of the focal groups that were affected and who confirmed the items were not biased before the items could be selected for use. The panelists did not have a vested interest in the outcome of the decision.

Psychometric review results, including the number of forms and number of items, are presented in table 4.1.

Table 4.1 Number of Forms and Items Reviewed Psychometrically

Grade Level or Grade Band	Number of Forms	Number of Unique Operational Items	Number of Unique Field Test Items	Total Number of Unique Items
Five	4	60	16	76
Eight	4	59	14	73
High School	4	59	15	74
Overall	12	178	45	223

Note: Some operational and field test items were administered across two or more forms.

4.3.3. Content Review of Forms

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

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

4.3.4. California Department of Education Forms Review

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

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

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

4.3.5. Configuration of the Test Delivery System

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

CAI's TDS supported a variety of item layouts. Some of the item layouts had the stimulus and item response options and response area displayed side by side. In each of these item layouts, both the stimulus and response options had independent scroll bars. Each item underwent an extensive platform review on different operating systems such as Windows, Linux, and iOS, to ensure that the item looked consistent across all platforms.

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

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

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

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

References

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx

California Department of Education. (2020). *California Alternate Assessment for Science*. Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/caascience.asp

Chapter 5: Test Administration

This chapter describes the administration of the embedded performance tasks (PTs) for the 2019–2020 California Alternate Assessment (CAA) for Science administration, as well as the procedures followed by ETS to ensure test security.

5.1. Student Test-Taking Requirements

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

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

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

5.2. Administration Preparations

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

5.2.1. Resources for the CAA for Science

Prior to the suspension of testing, to ensure the 2019–2020 test administration was a successful experience for CAA test examiners and students, ETS provided an online, self-guided training tutorial for CAA test examiners (2019a) as well as on-site test administration workshops in various locations throughout California in January 2020. ETS also produced webcasts and videos for detailed information on California Assessment of Student Performance and Progress (CAASPP) test administration procedures. The on-site workshops included a session dedicated exclusively to the topic of the CAA test administration procedures.

In addition, ETS developed and posted a number of test administration resources for schools and LEAs on both the public website at https://www.caaspp.org/ and on the secure Test Operations Management System (TOMS) website. These resources included detailed information on topics such as technology readiness, test administration, test security, accessibility resources, using the test delivery system (TDS), and general testing rules.

Given that the CAAs are administered to students who have the most significant cognitive disabilities, a test examiner—usually the student's teacher, who is familiar to the student—administers the CAA to the student one on one.

5.2.2. Practice and Training Tests

The publicly available practice and training tests are provided to prepare students for the summative assessment. These tests, available for grades five and eight and high school, simulate the experience of the CAA for Science online assessments. Practice and training tests align with Science Connectors but do not produce scores. Students may access them using a web browser.

The purposes of the practice and training tests are to

- allow students and administrators to become familiar with the user interface and components of the test delivery system (TDS) and the process of starting and completing a testing session; and
- introduce students and administrators to new grade-specific items similar to those on the operational assessment.

5.2.3. Local Educational Agency Training

Each year, ETS, in collaboration with the California Department of Education (CDE) and their Assessment Validity and Outreach contractor, the Sacramento County Office of Education (SCOE), establishes and implements a comprehensive training plan for LEA assessment staff and educators on all aspects of the assessment program. The ETS and SCOE annual training plans are developed with stakeholder feedback and specify the audience, topics, frequency, and mode (in-person, webcast, videos, modules, etc.) of the training, including such elements as format, participants, and logistics.

In 2019−2020 ETS and SCOE increased their collaboration efforts to provide a more streamlined training experience for LEA and school staff. ETS and SCOE began coordinating training plans and posting all training opportunities in one centralized location on the CAASPP website. LEA staff were able to register for training opportunities across both CDE contractors' offerings in one place and access all archived training materials on the 2019−2020 Training Opportunities web page. This new streamlined and coordinated process provides easy access to all the trainings that were offered.

5.2.3.1. Workshops

ETS conducted eight in-person pretest workshops and a pretest webcast in January 2020 for the 2019–2020 administration, which focused on training LEA CAASPP coordinators on how to prepare for administering all aspects of the CAASPP online assessments.

Training was also provided to focus on interpreting and using results. ETS typically provides eight in-person "CAASPP Results Are In—Now What?" workshops. However, because of the impact of the novel coronavirus disease 2019 pandemic on statewide testing, this workshop was converted to a virtual training. With the cancellation of statewide testing, and limited results being released to LEAs, the title of the training was changed to "CAASPP: Using Assessment Data for Decision-Making." This training was made available to LEAs as four stand-alone modules that focused on what data can tell educators about current student learning, how to interpret data, how to communicate data to local stakeholders, and making sense of Smarter Balanced data.

In addition to the in-person training opportunities offered by ETS, SCOE held the first California Assessment Conference (CAC) in Oakland, California, in October 2019. This conference focused on building connections between assessments and the classroom by

providing classroom educators with information about using statewide assessment data to improve teaching and learning. The conference included sessions on the three parts of a balanced assessment system: formative assessments, interim assessments, and summative assessments. The conference also offered sessions that provided information about accessibility resources.

5.2.3.2. Virtual Training and Webcasts

ETS provided a series of virtual trainings and live webcasts throughout the school year that were archived and made available for training LEA and test site staff as well as test administrators and test examiners. Webcast viewers were provided with a method of electronically submitting questions to the presenters during the webcast. The webcasts were recorded and archived for on-demand viewing on the 2019−2020 CAASPP Archived Webcasts web page at https://bit.ly/3xZRZjF. CAASPP webcasts were available to everyone and required neither preregistration nor a logon account.

In addition to the webcasts provided by ETS, SCOE hosted a number or virtual trainings intended to support new LEA CAASPP coordinators throughout all aspects of administration. This training series provided opportunities for new LEA CAASPP coordinators to receive timely training nearly every month.

SCOE also offered assessment update meetings intended to provide LEA CAASPP coordinators with timely updates about California's assessment system. The meetings were recorded and archived.

5.2.3.3. Videos

To supplement the virtual trainings, webcasts, and in-person workshops, ETS also produced short demonstration videos on various aspects of administering the CAASPP, which were available on the CAASPP Quick Reference Guides and Videos web page at https://www.caaspp.org/administration/instructions/qrgs-and-videos/index.html. SCOE produced quick reference guides to accompany many of the video resources, providing multiple avenues of support for educators administering the assessments.

5.2.3.4. Training for Proper Assignment of Designated Supports and Accommodations

ETS produced short demonstration videos for every embedded accessibility resource that demonstrated how to use the resource for educators, students, and parents. The videos were available in both English and Spanish on the Accessibility Resources Demonstration Videos web page at https://www.caaspp.org/training/caaspp/uaag.html.

In addition, ETS developed a video with LEA staff about the importance of implementing CAASPP accessibility resources with to help California educators learn more about the importance of accessibility resources and best practices used by educators in the field. The "Importance of Implementing CAASPP and ELPAC Accessibility Resources: Voices from Educators" video was available on the CAASPP Training Video web page at https://www.caaspp.org/rsc/videos/archived-training_importance-of-implementing-accessibility.html.

A video on how to use the Individual Student Assessment Accessibility Profile (ISAAP) Tool was also available to support educators in the process of creating an individual student profile and matching accessibility resources to student needs to ensure a fair and valid testing experience.

At the CAC, SCOE offered three sessions on accessibility. A Plenary Accessibility 101 session was presented to all conference attendees, intended to build a shared understanding of basic accessibility-related terms and considerations. The Creating an Equitable Process breakout session focused on developing an equitable and systematic process for matching students with appropriate accessibility resources. Matching Resources to Student Needs was another breakout session focused on providing an opportunity to practice appropriately matching student needs to the various accessibility resources.

5.3. Administration of the Embedded Performance Tasks

The CAA for Science operational assessment was administered one-on-one by a test examiner familiar with the student being tested. The test examiner administered four embedded PTs to each student; these were administered online through the CAASPP TDS.

5.3.1. Administration of Orienting Activities

Each embedded PT has two orienting activities, one for each of the two Connector sets in an embedded PT. The orienting activities were administered one-one-one by the test examiner prior to presenting the first item in each Connector set to the student. The administration of the items in each Connector set in an embedded PT should directly follow the delivery of each orienting activity.

5.3.2. Administration of the Embedded Performance Tasks

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.

5.3.3. Administration of the Test Administration Survey

During the 2019–2020 administration year, test examiners were asked to respond to a survey about their students. After the embedded PT was administered to the student, test examiners were presented with two surveys, with the instruction to respond to only one of the surveys on the basis of whether or not the student had been responsive during the testing session. The purpose of the survey was to collect basic information about students' experiences with the assessment process.

The surveys were included in the last section of each embedded PT delivered through the TDS.

5.4. Procedures to Maintain Standardization

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

5.4.1. Local Educational Agency CAASPP Coordinator

An LEA CAASPP coordinator was designated by the district superintendent at the beginning of the 2019–2020 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 TOMS;
- training CAASPP test site coordinators and test examiners regarding state requirements and CAA administration, as well as security policies and procedures;
- reporting test security incidents (including testing irregularities) to the CDE;
- overseeing test administration activities;
- filing a report of a testing incident in the Security and Test Administration Incident Reporting System (STAIRS); and
- requesting an Appeal (if indicated by TOMS prompts while reporting an incident using the STAIRS/Appeals process).

5.4.2. 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 CAASPP test site coordinator is responsible for identifying test examiners and ensuring that they have signed CAASPP Test Security Affidavits (5 *CCR* Section 859[d]). CAASPP test site coordinators' duties may include

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

5.4.3. 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 CAA for Science;
- reporting all test security incidents to the test site coordinator and LEA CAASPP coordinator in a manner consistent with state and LEA policies;
- viewing student information prior to testing to ensure that the correct student receives the proper test with appropriate resources and reporting potential data errors to test site coordinators and LEA CAASPP coordinators;
- monitoring student progress throughout the test session using the Test Administrator Interface; and
- complying fully with all directions provided in the *Directions for Administration (DFA)* for the CAA for Science.

5.4.4. Instructions for Test Examiners and Staff Involved in CAA for Science Administration

5.4.4.1. *Directions for Administration*

Test examiners used the embedded PT *DFAs* 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, the exemplar script, and possible individualization. *DFAs* also included scoring rubrics where warranted. Each of the four embedded PTs for a grade level is administered using a separate *DFA*.

Sample *DFAs* for the CAAs to be used in conjunction with the CAA practice and training tests were provided to LEAs as well (CDE, 2019c).

5.4.4.2. *CAASPP Online Test Administration Manual*

The *CAASPP Online Test Administration Manual* (CDE, 2020d) contains information and instructions on overall procedures and guidelines for all LEA and test site staff involved in the administration of online assessments as well as for the CAA for Science. Sections included the following topics:

- Roles and responsibilities of those involved with CAASPP testing
- Test administration resources
- Test security
- Administration preparation and planning
- General test administration
- Instructions for steps to take before, during, and after testing

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

5.4.4.3. *CAASPP and English Language Proficiency Assessments for California (ELPAC) Test Operations Management System (TOMS) User Guide*

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

TOMS modules described in the *TOMS User Guide* included the following (CDE, 2020c):

- **Test Administration Setup—**This module allowed LEAs to determine and calculate dates for the LEA's 2019–2020 testing.
- **Adding and Managing Users—**This module allowed LEA CAASPP coordinators to add CAASPP test site coordinators and test examiners to TOMS so that the designated user could access the online embedded PT *DFAs*.
- **Student Test Assignment—**This module allowed LEA CAASPP coordinators to designate students to take the alternate assessments.

5.4.4.4. CAA for Science Administration Planning Guides

The administration planning guides, posted prior to the annual launch of the embedded PTs, provided information about the embedded PTs that will be administered in the coming school year (CDE, 2020a). The administration planning guides contained information to help test examiners understand how to plan for the administration of the embedded PTs throughout the school year, version assignments, and test security. The administration planning guides also contained the following information:

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

5.4.4.5. Other System Manuals

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

- ***CAASPP and ELPAC Technical Specifications and Configuration Guide for Online Testing—***This manual 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, 2019b).
- ***CAASPP Security Incidents and Appeals Procedure Guide—***This manual provides information on how to report a testing incident and submit an Appeal to the CDE to reset, reopen, invalidate, or restore individual online student assessments (CDE, 2020e).
- ***CAASPP and ELPAC Accessibility Guide for Online Testing—***This manual provides descriptions of the accessibility features for online tests as well as information about supported hardware and software requirements for administering tests to students using accessibility resources, including those with a braille accommodation using Job Access With Speech (JAWS®) (software) or a braille embosser (hardware) (CDE, 2020b).

5.5. Accessibility Features for the 2019–2020 Administration

5.5.1. Individualizations

A notable feature of the 2019–2020 embedded PTs is that test examiners had the option to individualize certain elements of the assessment, although not all embedded PTs allowed for individualization. For the operational assessment administration, test examiners were instructed to review the activities associated with each embedded PT and decide whether the exemplar activity met a student's needs or if an individualized activity was appropriate. The test examiner documented the use of individualizations in the survey at the end of each embedded PT.

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

Table 5.1 through table 5.3 display the results of the survey regarding the kinds of individualization provided. The n-counts in these tables are based on all students in the statistical analysis file. Although test examiners were permitted to individualize the administration of the CAA for Science, table 5.1 through table 5.3 indicate that few students received individualizations, meaning the majority of students were administered the embedded PTs as outlined in the *DFA*s.

Table 5.1 Individualizations—Grade Five

Individualization	PT 1 (Life Sciences), Activity 1—Number	PT 1 (Life Sciences), Activity 1—Percent of Total	PT 1 (Life Sciences), Activity 2—Number	PT 1 (Life Sciences), Activity 2—Percent of Total	PT 2 (Physical Sciences), Activity 1—Number	PT 2 (Physical Sciences), Activity 1—Percent of Total	PT 2 (Physical Sciences), Activity 2—Number	PT 2 (Physical Sciences), Activity 2—Percent of Total	PT 3 (Earth and Space Sciences), Activity 1—Number	PT 3 (Earth and Space Sciences), Activity 1—Percent of Total	PT 3 (Earth and Space Sciences), Activity 2—Number	PT 3 (Earth and Space Sciences), Activity 2—Percent of Total
Using Standardized Scripts	732	99%	733	99%	735	100%	736	100%	730	99%	733	99%
Using Individualized Scripts	5	1%	4	1%	2	0%	1	0%	7	1%	4	1%
Using Standardized Materials	728	99%	730	99%	722	98%	731	99%	727	99%	731	99%
Using Individualized Materials	9	1%	7	1%	15	2%	6	1%	10	1%	6	1%

Table 5.2 Individualizations—Grade Eight

Individualization	PT 1 (Life Sciences), Activity 1—Number	PT 1 (Life Sciences), Activity 1—Percent of Total	PT 1 (Life Sciences), Activity 2—Number	PT 1 (Life Sciences), Activity 2—Percent of Total	PT 2 (Physical Sciences), Activity 1—Number	PT 2 (Physical Sciences), Activity 1—Percent of Total	PT 2 (Physical Sciences), Activity 2—Number	PT 2 (Physical Sciences), Activity 2—Percent of Total	PT 3 (Earth and Space Sciences), Activity 1—Number	PT 3 (Earth and Space Sciences), Activity 1—Percent of Total	PT 3 (Earth and Space Sciences), Activity 2—Number	PT 3 (Earth and Space Sciences), Activity 2—Percent of Total
Using Standardized Scripts	799	99%	798	99%	792	99%	795	99%	792	99%	794	99%
Using Individualized Scripts	5	1%	6	1%	12	1%	9	1%	12	1%	10	1%
Using Standardized Materials	791	98%	797	99%	788	98%	792	99%	797	99%	801	100%
Using Individualized Materials	13	2%	7	1%	16	2%	12	1%	7	1%	3	0%

Table 5.3 Individualizations—High School

Individualization	PT 1 (Life Sciences), Activity 1—Number	PT 1 (Life Sciences), Activity 1—Percent of Total	PT 1 (Life Sciences), Activity 2—Number	PT 1 (Life Sciences), Activity 2—Percent of Total	PT 2 (Physical Sciences), Activity 1—Number	PT 2 (Physical Sciences), Activity 1—Percent of Total	PT 2 (Physical Sciences), Activity 2—Number	PT 2 (Physical Sciences), Activity 2—Percent of Total	PT 3 (Earth and Space Sciences), Activity 1—Number	PT 3 (Earth and Space Sciences), Activity 1—Percent of Total	PT 3 (Earth and Space Sciences), Activity 2—Number	PT 3 (Earth and Space Sciences), Activity 2—Percent of Total
Using Standardized Scripts	712	98%	721	99%	722	99%	725	99%	715	98%	719	99%
Using Individualized Scripts	17	2%	8	1%	7	1%	4	1%	14	2%	10	1%
Using Standardized Materials	727	100%	727	100%	724	99%	724	99%	704	97%	728	100%
Using Individualized Materials	2	0%	2	0%	5	1%	5	1%	25	3%	1	0%

5.5.2. Type and Level of Accommodations

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

5.6. Processing and Scoring

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

All item types were designed to be machine-scorable.

5.7. Test Security and Confidentiality

5.7.1. ETS' Office of Testing Integrity

The Office of Testing Integrity (OTI) is a division of ETS that provides quality assurance services for all ETS-managed testing programs. This division resides in the ETS legal department. The Office of Professional Standards Compliance at ETS publishes and maintains the *ETS Standards for Quality and Fairness* (2014), which supports the OTI's goals and activities. The *ETS Standards for Quality and Fairness* provides guidelines to help ETS staff design, develop, and deliver technically sound, fair, and beneficial products and services and help the public and auditors evaluate those products and services.

The OTI's mission is to

- minimize any testing security violations that can impact the fairness of testing,
- minimize and investigate any security breach that threatens the validity of the interpretation of test scores, and
- report on security activities.

The OTI helps prevent misconduct on the part of students and administrators, detects potential misconduct through empirically established indicators, and resolves situations involving misconduct in a fair and balanced way that reflects the laws and professional standards governing the integrity of testing. In its pursuit of enforcing secure testing practices, the OTI strives to safeguard the various processes involved in a test development and administration cycle. For the CAA for Science, those processes included the following:

- Security of electronic files using a firewall
- Printing and publishing
- Test administration
- Test delivery
- Processing and scoring

- Data management
- Statistical analysis
- Student confidentiality

5.7.2. 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 *DFAs* that contain secure information from the embedded PTs (CDE, 2020e).

The following measures ensured the security of CAASPP System assessments administered in 2019–2020:

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

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

5.7.3. Security of Electronic Files Using a Firewall

A firewall is software that prevents unauthorized entry to files, email, and other organization-specific information. All ETS data exchanges and internal email remain within the ETS firewall at all ETS locations, ranging from Princeton, New Jersey; to San Antonio, Texas; to Concord and Sacramento, California.

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

Refer to section *1.9 Systems Overview and Functionality* in *Chapter 1: Introduction* for more information on TOMS.

5.7.4. Transfer of Scores via Secure Data Exchange

Because of the confidential nature of test results, ETS currently uses secure file transfer protocol (SFTP) and encryption for all data file transfers; test data is never sent via email. SFTP is a method for reliable and exclusive routing of files. Files reside on a password-protected server that only authorized users can access. ETS shares an SFTP server with the CDE. On that site, ETS posts Microsoft Word and Excel files, Adobe Acrobat PDFs, or other document files for the CDE to review; the CDE returns reviewed materials in the same manner. Files are deleted upon retrieval.

The SFTP server is used as a conduit for the transfer of files; secure test data is 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.

5.7.5. 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.

5.7.6. Statistical Analysis on Secure Servers

During all CAASPP testing, ETS information technology staff members retrieve data files from Cambium Assessment, Inc. (CAI) and load those files into a database. The ETS Data Quality Services staff extracts the data from the database and performs quality control procedures (e.g., the values of all variables are as expected) before passing files to the ETS statistical analysis group. The statistical analysis staff store the files on secure servers. All staff members involved with the data adhere to the ETS Code of Ethics and the ETS Information Protection Policies to prevent any unauthorized access to data.

5.7.7. 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.

5.7.8. Security and Test Administration Incident Reporting System Process

Test security incidents, such as improprieties, irregularities, and breaches, are prohibited behaviors that give a student an unfair advantage or compromise the secure administration of the tests, which, in turn, compromises the reliability and validity of test results (CDE, 2020e). Whether intentional or unintentional, failure by staff or students to comply with security rules constitutes a test security incident. Test security incidents have impacts on scoring and affect students' performance on the test.

LEA CAASPP coordinators and CAASPP test site coordinators must ensure that all test security and summative administration incidents are documented by following the prompts in TOMS that guided coordinators in their submittal. An Appeal is a request to reset, restore, reopen, invalidate, or grant a grace period extension to a student's test. If an Appeal to a student's test was warranted, TOMS provided additional prompts to file the Appeal.

After the form was submitted, an email containing a case number and next steps was sent to the submitter (and to the LEA CAASPP coordinator, if the case was submitted by the CAASPP test site coordinator). The STAIRS case in TOMS provided the LEA CAASPP coordinator, the CDE, and the California Technical Assistance Center (CalTAC) with the opportunity to interact and communicate regarding the STAIRS process (CDE, 2020e).

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

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

Appeals requests were reviewed by the CDE. When a request to submit an Appeal was approved, the coordinator received a system-generated email with the Appeal type that was approved (CDE, 2020e).

5.7.8.1. 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 should report the incident within 24 hours, using the online STAIRS/Appeals process in TOMS.

5.7.8.2. 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 and submitted using the online STAIRS/Appeals process in TOMS. An irregularity must be reported to the LEA CAASPP coordinator and CAASPP test site

coordinator immediately. The coordinator must report the irregularity within 24 hours, using the online STAIRS/Appeals process in TOMS.

5.7.8.3. 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 report the incident using the online STAIRS/Appeals process in TOMS within 24 hours. Examples may include such situations as a release of secure materials or a security or system risk. These circumstances have external implications for the CDE and may result in a decision to remove the test item(s) from the available secure item bank. A breach incident must be reported to the LEA CAASPP coordinator immediately.

5.7.9. Appeals

For test security incidents reported in STAIRS that resulted in a need to reset, reopen, invalidate, or restore individual online student assessments, the request was approved by the CDE. In most instances, an Appeal was submitted to address a test security breach or irregularity. The LEA CAASPP coordinator or CAASPP test site coordinator submitted Appeals in TOMS. All submitted Appeals are available for retrieval and review by the appropriate credentialed users within a given organization. However, the view of Appeals is restricted according to the user role as established in TOMS. An Appeal could be requested only by the LEA CAASPP coordinator or CAASPP test site coordinator if prompted while filing a STAIRS case in TOMS (CDE, 2020e).

Types of Appeals available during the 2019–2020 CAASPP administration are described in table 5.4.

Table 5.4 Types of Appeals in CAASPP Testing

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

Table 5.5 presents the number of Appeals in STAIRS in the 2019–2020 administration for each grade level or grade band.

Table 5.5 Number of Appeals in STAIRS in the 2019–2020 Administration—All Grades

Appeal Type	Total Count of Appeals
Reset	1
Invalidate	0
Re-open	2
Restore	0
Grace Period Extension	0
Swap Approved	0
No Appeal	9

Table 5.6 presents the number of testing issues reported in STAIRS by type.

Table 5.6 Number of Testing Issues Reported in STAIRS by Type

Testing Issue	All Grades
Accessibility Issue	2
Accidental Summative Access	0
Administered Incorrect Assessment	0
Administration Error	8
Data Entry Issue	0
Disruption or Technical Issues	2
Domain Exemptions or Incorrect Alternate Assessment	0
Expired or Accidentally Submitted Test	2
Exposing Secure Materials	0
Incorrect SSID Used	0
Restore from Reset	0
Student Cheating or Accessing Unauthorized Devices	0
Student Disruption	0
Other Issues	0

5.8. Monitoring Assessment of Students

The CAA for Science 2019–2020 operational assessment offered commonly used accessibility resources available through the CAASPP online testing platform, where applicable for the tested construct.

5.8.1. Universal Tools and Designated Supports for Students with Disabilities

The purpose of universal tools, designated supports, and accommodations in testing is to allow *all* students the opportunity to demonstrate what they know and what they are able to do, rather than giving students who use these resources an advantage over other students or artificially inflating their scores. Universal tools, designated supports, and accommodations minimize or remove barriers that could otherwise prevent students from demonstrating their knowledge, skills, and achievement in a specific content area.

5.8.1.1. Universal Tools

Universal tools are available to all students by default, although they can be disabled if a student finds them distracting. Each universal tool falls into one of two categories: embedded and non-embedded. Embedded universal tools are provided through the student testing interface (through the CAASPP secure browser), although they can be turned off by a test administrator.

The resources in the following subsections were available in the 2019–2020 CAA for Science administration.

5.8.1.1.1. Embedded

- Breaks
- Digital notepad
- Expandable items
- Expandable passages
- Highlighter
- Keyboard navigation
- Line reader
- Mark for review
- Strikethrough
- Writing tools (e.g., bold, italic, bullets, undo or redo) (for specific items)
- Zoom (in or out)

5.8.1.1.2. Non-embedded

- Breaks
- Scratch paper

5.8.1.2. Designated Supports

Designated supports are available to all students through the test settings in TOMS. The designated supports each fall into one of two categories: embedded and non-embedded. Embedded designated supports are provided through the student testing interface (through the CAASPP secure browser).

The resources in the following subsections were available in the 2019–2020 CAA for Science administration.

5.8.1.2.1. Embedded

- Color contrast
- Masking
- Mouse pointer (size and color)

- Permissive mode
- Streamline
- Turn off any universal tool(s)

5.8.1.2.2. Non-embedded

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

5.8.1.3. Accommodations

Accommodations are changes in procedures or materials that increase equitable access during the CAASPP assessments. Assessment accommodations generate valid assessment results for students who need them; they allow these students to show what they know and can do. Accommodations do not compromise the learning expectations, construct, grade-level standard, or intended outcome of the assessments.

The resources in the following subsections were available in the 2019–2020 CAA for Science administration. Note there were no embedded accommodations for the CAA for Science.

5.8.1.3.1. Non-embedded

- Abacus
- Additional instructions supports and resources for alternate assessments
- Alternate response options
- Print on demand
- Word prediction

5.8.2. Identification

All public school students participate in the CAASPP System, including students with disabilities and English learner students. The *Smarter Balanced Assessment Consortium: Usability, Accessibility, and Accommodations Guidelines* (Smarter Balanced, 2020) and the CDE's Matrix One (CDE, 2019d) are intended for school-level personnel and IEP and Section 504 plan teams to select and administer the appropriate universal tools, designated supports, and accommodations as deemed necessary for individual students.[7] The CAA for Science assessments follow the Smarter Balanced recommendations for use (Smarter Balanced, 2020).

The *Guidelines* apply to all students and promote an individualized approach to the implementation of assessment practices. Another web document, the *Smarter Balanced*

[7] This technical report is based on the version of Matrix One that was available during the 2019–2020 CAASPP administration. Note that Matrix One has since been combined with the English Language Proficiency Assessments for California Matrix Four to form a single

Resources and Practices Comparison Crosswalk (Smarter Balanced, 2018), connects the assessment resources described in the *Guidelines* with associated classroom practices.

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

5.8.3. Assignment

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

There are three ways the student's accessibility resource(s) could be assigned:

1. Using the ISAAP Tool to identify the accessibility resource(s) and then uploading the spreadsheet it creates into TOMS (This process is discussed in more detail in subsection *2.5.1.1 Resources for Selection of Accessibility Resources*.)
2. Using the Online Student Test Settings template to enter students' assignments and then uploading the spreadsheet into TOMS
3. Entering assignments for each student individually in TOMS

If a student's IEP or Section 504 plan team identified and designated a resource not identified in an accessibility matrix, the LEA CAASPP coordinator or CAASPP test site coordinator needed to submit a request for an unlisted resource to be approved by the CDE. The CDE then determined whether the requested unlisted resource changed the construct being measured after all testing was completed.

5.8.4. Accommodation Usage

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

Table 5.7 through table 5.9 report the number of students who, based on the availability of data, were assigned to a certain accommodation or designated support and actually used this accommodation or designated support.

Types of accommodations and designated supports—labeled "ACC" and "DS" in the *Resource Type* column—included in table 5.7 through table 5.9 are listed as follows:

- **Print on Demand:** Paper copies of passages and stimuli, items, or all of these are printed for students.
- **Masking:** This resource involves blocking off content that is not of immediate need or that may be distracting to the student.

accessibility resources matrix, the California Assessment Accessibility Resources Matrix (CDE, 2020f).

The other types of accommodations and designated supports were not used during the 2019–2020 CAA for Science administration and, therefore, are not included in these tables.

Table 5.7 Summary of Accommodations and Designated Supports Used by Students, Grade Five

Resource Type	# of Students Assigned an ACC or DS for ESS	# of Students Who Used an ACC or DS for ESS	% of Students Assigned an ACC or DS for PS	% of Students Who Used an ACC or DS for PS	% of Students Assigned an ACC or DS for LS	% of Students Who Used an ACC or DS for LS
ACC—Non-Embedded Print on Demand	2	1	2	0	0	0
DS—Embedded Masking	3	0	17	1	7	0

Note: ESS = Earth and Space Sciences, PS = Physical Sciences, and LS = Life Sciences.

Table 5.8 Summary of Accommodations and Designated Supports Used by Students, Grade Eight

Resource Type	# of Students Assigned an ACC or DS for ESS	# of Students Who Used an ACC or DS for ESS	% of Students Assigned an ACC or DS for PS	% of Students Who Used an ACC or DS for PS	% of Students Assigned an ACC or DS for LS	% of Students Who Used an ACC or DS for LS
ACC—Non-Embedded Print on Demand	15	1	8	1	2	1
DS—Embedded Masking	34	0	24	0	6	0

Note: ESS = Earth and Space Sciences, PS = Physical Sciences, and LS = Life Sciences.

Table 5.9 Summary of Accommodations and Designated Supports Used by Students, High School

Resource Type	# of Students Assigned an ACC or DS for ESS	# of Students Who Used an ACC or DS for ESS	% of Students Assigned an ACC or DS for PS	% of Students Who Used an ACC or DS for PS	% of Students Assigned an ACC or DS for LS	% of Students Who Used an ACC or DS for LS
ACC—Non-Embedded Print on Demand	0	0	0	0	0	0
DS—Embedded Masking	12	0	9	0	3	0

Note: ESS = Earth and Space Sciences, PS = Physical Sciences, and LS = Life Sciences.

References

California Department of Education. (2019a). *2019–20 California Alternate Assessment for Science test examiner tutorial*. Sacramento, CA: California Department of Education. https://www.brainshark.com/1/player/ets?fb=0&r3f1=&custom=caa_test_examiner_tutorials

California Department of Education. (2019b). *CAASPP and ELPAC technical specifications and configuration guide for online testing*. Sacramento, CA: California Department of Education. https://bit.ly/3upqiyD

California Department of Education. (2019c). *California Alternate Assessment for Science directions for administration, training performance task, fossils and plate tectonics.* Sacramento, CA: California Department of Education. https://bit.ly/2Q3HdI9

California Department of Education. (2019d). *Matrix one: Universal tools, designated supports, and accommodations for the California Assessment of Student Performance and Progress for 2019–20.* Sacramento, CA: California Department of Education. https://bit.ly/3h0EkD2

California Department of Education. (2020a). *2020–21 CAA for Science administration planning guide: Grade eight.* Sacramento, CA: California Department of Education. https://bit.ly/3o3nepq

California Department of Education. (2020b) *CAASPP and ELPAC accessibility guide for online testing.* Sacramento, CA: California Department of Education. https://bit.ly/3usOzny

California Department of Education. (2020c). *CAASPP and ELPAC Test Operations Management System user guide*. Sacramento, CA: California Department of Education. https://bit.ly/2QKKJaV

California Department of Education. (2020d). *CAASPP online test administration manual, 2019–2020 test administration.* Sacramento, CA: California Department of Education. https://bit.ly/3eXYBXc

California Department of Education. (2020e). *CAASPP security incidents and Appeals procedure guide.* Sacramento, CA: California Department of Education. https://bit.ly/3b5saoH

California Department of Education. (2020f). *California assessment accessibility resources matrix*. Sacramento, CA: California Department of Education. https://bit.ly/3vHmKbm

California Science Teachers Association. (2000–2019). *CA Alternate Assessment - CAA-Science*. https://cascience.org/ngss/assessments/ca-alternate-assessment-caa-science

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

Smarter Balanced Assessment Consortium. (2014). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations implementation guide.* Los Angeles: Smarter Balanced Assessment Consortium. https://bit.ly/3xN8U9b

Smarter Balanced Assessment Consortium. (2018). *Smarter Balanced Resources and Practices Comparison Crosswalk.* Los Angeles: Smarter Balanced Assessment Consortium. https://bit.ly/3h5xBHS

Smarter Balanced Assessment Consortium. (2020). *Smarter Balanced Assessment Consortium: Usability, accessibility, and accommodations guidelines.* Los Angeles, CA: Smarter Balanced Assessment Consortium. https://portal.smarterbalanced.org/library/en/usability-accessibility-and-accommodations-guidelines.pdf

Chapter 6: Scoring and Reporting

Student scores for the operational administration of the California Alternate Assessment (CAA) for Science, given during the 2019–2020 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 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.

6.1. CAA for Science Scoring Process

Each student was administered three operational embedded performance tasks (PTs) and one field test embedded PT, each consisting of 10 items, for a total of 12 points. Two items in each embedded PT are worth two points. The field test items do not count toward the student's total test score.

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

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

6.2. Types of Scores

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

6.2.1. Percent Correct

The percent-correct scores are calculated for operational items. The percent correct is calculated using the following equation:

$$Percent\ correct = \frac{Number\ of\ points\ earned\ for\ all\ items}{Maximum\ number\ of\ points\ possible\ for\ all\ items} \quad (6.1)$$

Refer to the Alternative Text for Equation 6.1 for a description of this equation.

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

6.2.2. 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 development of the operational reporting scale.

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

Table 6.1 Indicator Categories

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

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

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

Table 6.2 Threshold Scores for Preliminary Categories

Grade Level or Grade Band	Required for Category 2	Required for Category 3
Grade 5	33%	90%
Grade 8	33%	90%
High school	33%	90%

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

Table 6.3 Grade Five Preliminary Indicator Conversion Table

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

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

Table 6.4 Grade Eight Preliminary Indicator Conversion Table

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

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

Table 6.5 High School Preliminary Indicator Conversion Table

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

Table 6.A.1 in appendix 6.A shows the percentage of students at each preliminary category level. Table 6.A.1 is based on all students who completed all four embedded PTs and who had a valid test score. The majority of students with scores at all grade levels were classified as being in preliminary category 2.

Reference

California Department of Education. (2019). *California Alternate Assessment for Science directions for administration, training performance task, fossils and plate tectonics.* Sacramento, CA: California Department of Education. https://bit.ly/2Q3HdI9

Accessibility Information

Alternative Text for Equation 6.1

Percent correct equals the number of points earned for all items divided by the maximum number of points for all items.

Appendix 6.A: Preliminary Indicator Summary

Table 6.A.1 Number and Percentage of Students in the Preliminary Indicator Categories

Grade Level or Grade Band	Preliminary Category 1 N	Preliminary Category 1 %	Preliminary Category 2 N	Preliminary Category 2 %	Preliminary Category 3 N	Preliminary Category 3 %	Total Number of Students
Five	57	16%	287	78%	23	6%	367
Eight	56	13%	351	81%	28	6%	435
HS—Ten	20	19%	78	73%	9	8%	107
HS—Eleven	17	8%	186	87%	12	6%	215
HS—Twelve	25	13%	149	78%	16	8%	190
HS—All Grades	62	12%	413	81%	37	7%	512

Chapter 7: Psychometric Analyses

This chapter summarizes the results of the psychometric analyses that are typically conducted for the California Alternate Assessment (CAA) for Science administration. However, because of the impact of the novel coronavirus disease 2019 (COVID-19) pandemic, no classical item analyses, differential item functioning (DIF) analyses, item response theory (IRT) analyses, reliability analyses, or validity analyses were conducted for the 2019–2020 CAA for Science administration.

7.1. Overview

This section describes the data samples typically used for the statistical analyses and provides explanations for all statistical procedures implemented in the psychometric analyses. Those procedures typically include item analyses, DIF analyses, IRT calibration, computation of reliability, and standard errors of measurement. The procedures are designed to ensure the validity of score uses.

7.1.1. Summary of the Analyses

ETS typically conducts the following analyses for the CAA for Science. However, these analyses were not conducted in 2019–2020 because of the impact of the COVID-19 pandemic. Each analysis is described in the body of this chapter.

1. **Classical Item Analyses—**Classical item analysis for the CAA for Science is discussed in section *7.2 Classical Item Analyses*.
2. **Omission and Completion Analyses—**The omit rate and item difficulty information for the CAA for Science are described in section *7.3 Omission and Completion Rates*.
3. **DIF Analyses—**DIF analysis for the CAA for Science is described in section *7.5 Differential Item Functioning Analyses*.
4. **IRT Analyses—**IRT calibration analyses for the CAA for Science are described in section *7.6 Item Response Theory Analyses*.
5. **Reliability Analyses—**Reliability estimation for the CAA for Science is illustrated in section *7.7 Reliability Analyses*.
6. **Validity Evidence—**Validity evidence related to the CAA for Science is discussed in section *7.8 Validity Evidence*.

7.1.2. Sample Used for the Analyses

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

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

Table 2.1 provides the number of students assigned to take the CAA for Science and the number of students who started the tests and completed all four embedded PTs. Except for grade ten, less than 20 percent of the registered students started the CAA for Science and even fewer students completed the assessment.

Because of the extremely small number of students who completed the CAA for Science when testing was suspended on March 18, 2020, no psychometric analyses were performed for the 2019–2020 administration. Instead, the types of analyses typically conducted are described here.

7.2. Classical Item Analyses

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

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

7.2.1. 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). Items with two response choices are flagged if their *p*-values are below 0.50, three-choice items are flagged if their *p*-values are below 0.30, and four-choice items are flagged if their *p*-values are below 0.20 (i.e., too difficult).

The formula for *p*-value for a dichotomous item is:

$$p - value_{dich} = \frac{\sum X_{ij}}{N_i} \quad (7.1)$$

Refer to the Alternative Text for Equation 7.1 for a description of this equation.

where,

X_{ij} is the score (1 or 0) 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.

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

$$p\text{–}value_{poly} = \frac{\sum X_{ij}}{N_i \times Max(X_i)}, \quad (7.2)$$

Refer to the Alternative Text for Equation 7.2 for a description of this equation.

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.

7.2.2. 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 operational embedded PTs. The estimation formula is:

$$r_{polyreg} = \frac{\hat{\beta} s_{tot}}{\sqrt{\hat{\beta}^2 s_{tot}^2 + 1}} \qquad (7.3)$$

Refer to the Alternative Text for Equation 7.3 for a description of this equation.

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).

For a polytomous item, there is a regression for each boundary between item scores, with all regressions for the same item sharing a common slope, β. For a polytomous item with k possible score values, there are k-1 regressions. Beta (β) is the common slope for all k-1 regressions.

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

7.2.3. 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 3 percent) of students obtaining any possible item score are flagged for further review. Such items may pose problems during the IRT calibrations. They need to be carefully reviewed and may need to be excluded from the item calibration analyses.

7.2.4. Summary of Classical Item Analysis Flagging Criteria

Items are flagged for review if the item analysis yields any of the six 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 0.50 for two-choice dichotomous items, 0.30 for three-choice dichotomous items, 0.20 for four-choice dichotomous items, or 0.30 for polytomous items.
3. Item-total correlation (polyserial) is below 0.20.
4. Among the highest-performing students (the top 20 percent), the number of students choosing any distractor is greater than the number of those choosing the key.
5. The omit rate is above 5 percent for dichotomous items or above 15 percent for polytomous items.
6. Any of the possible scores on a polytomous item is earned by less than 3 percent of the students.

In a typical administration, ETS' psychometric staff and content assessment development staff review each of the flagged items and summarize the classical item results for the California Department of Education (CDE), with recommendations for subsequent analyses of the items. The classical item statistics are entered into the item bank for use by the assessment development team for test assembly for future operational administrations.

7.3. Omission and Completion Rates

7.3.1. Omit Rates

For both dichotomous and polytomous items, examining the omit rate is useful for identifying potential problems with test features such as testing time and item or test layout. An item is considered "omitted" when the item has been presented to the student but has not been answered (i.e., left blank) in the middle of an administered assessment wherein the student has been presented with, and responded to, successive items.

The *Mark as No Response* option is a specific case of an omitted item. The *Mark as No Response* option should be used when the item was presented to the student and the student did not provide a response despite the test examiner's best efforts to elicit a response. Similar to the omit rate, the *Mark as No Response* information is useful for identifying potential problems with an item.

7.3.2. Completion Rates

Completion rates indicate the proportion of students who completed each of the three embedded PTs on the test. A student's record for the CAA for Science is not considered complete unless the student answered at least one test question from each of the four embedded PTs.

7.4. Task Difficulty (Overall and by Embedded Performance Task)

The mean raw score for each embedded PT is provided in table 7.1 through table 7.3 and is based on the students in the preliminary indicator file that had a valid test score. Because testing was suspended on March 18, 2020, and so few students completed the CAA for Science, no interpretation of the results is provided.

Table 7.1 Raw Score Summary for Each Embedded PT—Grade Five

Module	Number of Students	Number of Items	Maximum Number of Points	Mean Raw Score	SD Raw Score	Minimum Raw Score	Maximum Raw Score	Mean Raw Score as a Percentage
PT 1 Version 1 (Life Sciences)	97	10	12	7.3	3.9	0	12	60.8
PT 2 Version 1 (Physical Sciences)	97	10	12	7.1	3.8	0	12	59.2
PT 3 Version 1 (Earth and Space Sciences)	97	10	12	6.2	3.5	0	12	51.7
PT 1 Version 2 (Life Sciences)	85	10	12	8.4	3.7	0	12	70.0
PT 2 Version 2 (Physical Sciences)	85	10	12	7.5	3.7	0	12	62.5
PT 3 Version 2 (Earth and Space Sciences)	85	10	12	6.6	3.1	0	12	55.0
PT 1 Version 3 (Life Sciences)	120	10	12	7.8	3.5	0	12	65.0
PT 2 Version 3 (Physical Sciences)	120	10	12	7.0	3.3	0	12	58.3
PT 3 Version 3 (Earth and Space Sciences)	120	10	12	7.7	3.6	0	12	64.2
PT 1 Version 4 (Life Sciences)	65	10	12	7.0	3.6	0	12	58.3
PT 2 Version 4 (Physical Sciences)	65	10	12	6.1	3.2	0	11	50.8
PT 3 Version 4 (Earth and Space Sciences)	65	10	12	6.7	3.6	0	12	55.8

Table 7.2 Raw Score Summary for Each Embedded PT—Grade Eight

Module	Number of Students	Number of Items	Maximum Number of Points	Mean Raw Score	SD Raw Score	Minimum Raw Score	Maximum Raw Score	Mean Raw Score as a Percentage
PT 1 Version 1 (Life Sciences)	111	10	12	7.0	3.3	0	12	58.3
PT 2 Version 1 (Physical Sciences)	111	10	12	7.4	3.7	0	12	61.7
PT 3 Version 1 (Earth and Space Sciences)	111	10	12	6.9	3.7	0	12	57.5
PT 1 Version 2 (Life Sciences)	95	10	12	6.2	2.9	0	12	51.7
PT 2 Version 2 (Physical Sciences)	95	10	12	7.0	3.3	0	12	58.3
PT 3 Version 2 (Earth and Space Sciences)	95	10	12	6.6	3.4	0	12	55.0
PT 1 Version 3 (Life Sciences)	147	10	12	7.2	2.9	0	12	60.0
PT 2 Version 3 (Physical Sciences)	147	10	12	7.6	3.3	0	12	63.3
PT 3 Version 3 (Earth and Space Sciences)	147	10	12	7.2	3.3	0	12	60.0
PT 1 Version 4 (Life Sciences)	82	10	12	6.6	3.2	0	12	55.0
PT 2 Version 4 (Physical Sciences)	82	10	12	7.1	3.6	0	12	59.2
PT 3 Version 4 (Earth and Space Sciences)	82	10	12	6.9	3.8	0	12	57.5

Table 7.3 Raw Score Summary for Each Embedded PT—High School

Module	Number of Students	Number of Items	Maximum Number of Points	Mean Raw Score	SD Raw Score	Minimum Raw Score	Maximum Raw Score	Mean Raw Score as a Percentage
PT 1 Version 1 (Life Sciences)	142	10	12	7.6	2.7	0	12	63.3
PT 2 Version 1 (Physical Sciences)	142	10	12	8.2	3.1	0	12	68.3
PT 3 Version 1 (Earth and Space Sciences)	142	10	12	7.3	2.7	0	12	60.8
PT 1 Version 2 (Life Sciences)	131	10	12	7.1	3.3	0	12	59.2
PT 2 Version 2 (Physical Sciences)	131	10	12	7.6	3.6	0	12	63.3
PT 3 Version 2 (Earth and Space Sciences)	131	10	12	6.4	3.4	0	12	53.3

Table 7.3 *(continuation)*

Module	Number of Students	Number of Items	Maximum Number of Points	Mean Raw Score	SD Raw Score	Minimum Raw Score	Maximum Raw Score	Mean Raw Score as a Percentage
PT 1 Version 3 (Life Sciences)	142	10	12	7.5	3.6	0	12	62.5
PT 2 Version 3 (Physical Sciences)	142	10	12	6.9	3.2	0	12	57.5
PT 3 Version 3 (Earth and Space Sciences)	142	10	12	6.9	3.3	0	12	57.5
PT 1 Version 4 (Life Sciences)	97	10	12	8.2	3.6	0	12	68.3
PT 2 Version 4 (Physical Sciences)	97	10	12	7.0	3.3	0	12	58.3
PT 3 Version 4 (Earth and Space Sciences)	97	10	12	7.6	3.5	0	12	63.3

7.5. Differential Item Functioning Analyses

In a typical CAA for Science administration, DIF analyses are conducted on the CAA for Science items with sufficient sample sizes. The minimum sample size requirements for the DIF analyses are 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, then the item may be measuring something other than the intended construct. It is important, however, to recognize that item performance differences flagged for DIF might be related to actual differences in relevant knowledge or skills between student groups (i.e., impact) or statistical Type I error, which might falsely find DIF in an item. As a result, DIF analysis is used mainly as a statistical tool to identify *potential* item bias. Subsequent reviews by content experts and bias and sensitivity experts are required to determine the source and meaning of performance differences.

7.5.1. Differential Item Functioning Procedure for Dichotomous Items

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

$$\alpha_{MH} = \frac{\left(\sum_m R_{rm} \frac{W_{fm}}{N_{tm}}\right)}{\left(\sum_m R_{fm} \frac{W_{rm}}{N_{tm}}\right)} \quad (7.4)$$

Refer to the Alternative Text for Equation 7.4 for a description of this equation.

where,

m indexes the score categories,

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

W_{rm} is the number of students in the reference 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_{fm} is the number of students in the focal group who answer the item incorrectly.

To facilitate the interpretation of MH results, the common odds ratio is transformed to the delta scale using the following formula (Holland & Thayer, 1988):

$$\mathrm{MH\ D\text{-}DIF} = -2.35 \ln\left[\alpha_{MH}\right] \quad (7.5)$$

Refer to the Alternative Text for Equation 7.5 for a description of this equation.

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

7.5.2. Differential Item Functioning Procedure for 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:

$$SMD = \frac{\sum_{m=1}^{M} N_{fm} \times (E_f(Y \mid X = m) - E_r(Y \mid X = m))}{\sum_{m=1}^{M} N_{fm}} \quad (7.6)$$

Refer to the Alternative Text for Equation 7.6 for a description of this equation.

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 in the reference group, and

E_f is the expected item score in the focal group.

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

7.5.3. Differential Item Functioning Categories and Definitions

Based on the DIF statistics and significance tests, items are classified into three categories, labeled A, B, and C. Category A items contain negligible DIF, Category B items exhibit slight to moderate DIF, and Category C items possess moderate to large DIF values.

The categorization rules for dichotomous items are presented in table 7.4; the categorization rules for polytomous items are presented in table 7.5.

Table 7.4 DIF Categories for Dichotomous Items

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

Table 7.5 DIF Categories for Polytomous Items

DIF Category	Criteria
A (negligible)	• Mantel Chi-square *p-value* > 0.05 or |SMD/SD| ≤ 0.17
B (moderate)	• Mantel Chi-square *p-value* < 0.05 and 0.17 < |SMD/SD| ≤ 0.25
C (large)	• Mantel Chi-square *p-value* < 0.05 and |SMD/*SD*| > 0.25

After a typical test administration, DIF analyses are conducted on each item for designated comparison groups, if the number of students in the group is sufficient. Groups are defined on the basis of demographic variables, including gender, race or ethnicity, and primary disabilities. These comparison groups are specified in table 7.6.

Table 7.6 Student Groups for DIF Comparison

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

Table 7.6 *(continuation)*

DIF Type	Reference Group	Focal Group
Intellectual Disability Versus Autism	Intellectual Disability Group, which includes the following: • Intellectual disability • Multiple disabilities • Traumatic brain injury	• Autism
Intellectual Disability Versus Learning Disability	Intellectual Disability Group, which includes the following: • Intellectual disability • Multiple disabilities • Traumatic brain injury	Learning Disability Group, which includes the following: • Emotional disturbance • Orthopedic impairment • Other health impairment • Specific learning disability • Speech or language impairment

* DIF analysis are not typically performed on these student groups because of insufficient sample sizes for the CAA for Science.

7.6. Item Response Theory Analyses

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

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

7.6.1. Item Response Theory Models

The one-parameter logistic (1PL) IRT model (1PL-IRT) is used for the CAA for Science item calibration and was selected after consultation with the CDE. The generalized partial credit model (GPCM) (Muraki, 1992) restricted for 1PL-IRT, which is essentially the partial credit model (PCM) (Masters, 1982), is applied to both dichotomous and polytomous items. The mathematical form of the GPCM is the following:

$$P_{ih}(\theta_j) = \begin{cases} \dfrac{\exp\left(\sum_{v=1}^{h} Da_i(\theta_j - b_i + d_{iv})\right)}{1 + \sum_{c=1}^{n_i} \exp\left(\sum_{v=1}^{c} Da_i(\theta_j - b_i + d_{iv})\right)}, & \text{if score } h = 1, 2, ..., n_i \\ \dfrac{1}{1 + \sum_{c=1}^{n_i} \exp\left(\sum_{v=1}^{c} Da_i(\theta_j - b_i + d_{iv})\right)}, & \text{if score } h = 0 \end{cases} \quad (7.7)$$

Refer to the Alternative Text for Equation 7.7 for a description of this equation.

where,

$P_{ih}(\theta_j)$ is the probability of student with proficiency θ_j obtaining score h on item i,

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

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

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

d_{iv} is the category difficulty parameter for item i on score v, and

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

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

7.6.2. Item Calibration

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

7.6.2.1. Data Preparation

Prior to IRT calibration analyses, ETS psychometricians review the results of the classical item analyses to decide whether any items are of poor quality and need to be removed from calibration. The results are also reviewed by ETS content experts and the CDE. Decisions to remove items from calibration are made in consultation with the CDE.

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

7.6.2.2. Description of the Calibration Procedure

FlexMIRT (Cai, 2016), a multilevel and multiple-group IRT software package for item analysis and test scoring, is used for CAA for Science item calibration analysis. This software can fit a variety of IRT models to both single-level and multilevel data that are dichotomous, polytomous, or both.

The calibration procedure is as follows:

1. Receive test form planners and create the item mapping files
2. Receive data

3. Run the complete classical item analyses
4. Prepare and format the sparse matrix input data files as required by flexMIRT
5. Prepare flexMIRT control files
6. Evaluate the flexMIRT output to examine whether every execution of flexMIRT analysis reached satisfactory convergence
7. Review the item parameter estimates:
    a. At the test level, the summary statistics for the *b*-parameter estimates (location difficulty) and *d*-parameter estimates (step difficulty) were examined, including the mean, SD, median, minimum, maximum, and model-fit. The model-fit was evaluated using the root mean square error of approximation (RMSEA). RMSEA values less than 0.05 indicate good fit while RMSEA values greater than 0.10 indicate poor fit (Browne & Cudeck, 1993). The *b*-parameters were correlated with the *p*-values.
    b. At the item level, statistics of individual items were examined, including item difficulty estimates (*b*-parameters and *d*-parameters) and item-fit statistics using the marginal chi-square statistic. The *b*-parameters and the *d*-parameters should be in the range of -4.0 to +4.0 with a standard error of 0.10 or less.
8. Flag items that did not perform as expected:
    a. All flagged items were discussed thoroughly with the CDE to decide whether those items should be removed from calibration or whether the scoring categories need to be collapsed.

The calibration process is conducted independently by two ETS psychometricians to ensure quality and accuracy of results. The two psychometricians independently create flexMIRT control files and run the same input data files and then compare the calibration results. Any differences in the output are investigated. Refer to section *9.4 Quality Control of Psychometric Processes* for more details of this procedure.

7.7. Reliability Analyses

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

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

There are several different ways of estimating reliability. Stratified Alpha is computed for the reliability estimates for each version of the CAA for Science for the student groups after a typical CAA administration. More details can be found in the next subsection, 7.7.1.

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

7.7.1. Internal Consistency Reliability

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

The formula for the internal consistency reliability as measured by Cronbach's Alpha (Cronbach, 1951) is:

$$\alpha = \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} \sigma_i^{\,2}}{\sigma_x^2}\right] \tag{7.8}$$

Refer to the Alternative Text for Equation 7.8 for a description of this equation.

where,

n is the number of items,

σ_i^2 is the variance of scores on the i-th item, and

σ_x^2 is the variance of the total score (sum of scores on the individual items).

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

The formula for calculating the stratified Alpha is:

$$\rho_{strat\alpha} = 1 - \frac{\sum \sigma_{X_j}^2 (1-\alpha_j)}{\sigma_X^2} \tag{7.9}$$

Refer to the Alternative Text for Equation 7.9 for a description of this equation.

where,

$\sigma_{X_j}^2$ is the variance for stratum j of the test,

σ_X^2 is the total variance of the test, and

α_j is the Cronbach's Alpha for stratum j of the test.

7.7.2. Standard Error of Measurement

The SEM provides a measure of score instability on the raw score metric. The SEM is the square root of the error variance in the scores (i.e., the SD of the distribution of the differences between students' observed scores and their true scores). The SEM is calculated by:

$$SEM = s_t\sqrt{1-\rho_{\hat{\theta}\hat{\theta}'}} \tag{7.10}$$

Refer to the Alternative Text for Equation 7.10 for a description of this equation.

where,

$\rho_{\hat{\theta}\hat{\theta}'}$ is the reliability estimated in equation 7.9, and

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

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

7.8. Validity Evidence

Validity refers to the degree to which each interpretation or use of a test score is supported by the accumulated evidence (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014; ETS, 2014). It constitutes the central notion underlying the development, administration, and scoring of tests and the uses and interpretations of test scores.

The validation process does not rely on a single study or gathering only one type of evidence. Rather, validation involves multiple investigations and different kinds of supporting evidence (AERA, APA, & NCME, 2014; Cronbach, 1971; ETS, 2014; Kane, 2006). It begins with the test design and is implicit throughout the entire assessment process, which includes item development and field testing, analyses of items, test scaling and linking, scoring, reporting, and score usage.

In this section, the evidence gathered is presented to support the intended uses and interpretations of scores for the CAA for Science. This section is organized primarily around the principles prescribed by AERA, APA, and NCME's *Standards for Educational and Psychological Testing* (2014). These *Standards* require a clear definition of the purpose of the test, a description of the constructs to be assessed, and the population to be assessed, as well as how the scores are to be interpreted and used.

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

1. Evidence based on test content
2. Evidence based on relations to other variables

3. Evidence based on response processes
4. Evidence based on internal structure
5. Evidence based on the consequences of testing

The next subsection defines the purpose of the CAA for Science, followed by a description and discussion of different kinds of validity evidence that are typically gathered.

7.8.1. Evidence in the Design of the CAA for Science

7.8.1.1. Purpose

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

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

7.8.1.2. The Constructs to Be Measured

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

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

- Subject to be assessed
- Tasks to be presented
- Administration instructions to be given
- Rules used to score student responses

The test blueprints control as many aspects of the measurement procedure as possible so that the testing conditions will remain the same over test administrations (Cronbach, 1971) to minimize construct-irrelevant score variance (Messick, 1989).

ETS developed all CAA for Science test items to conform to the State Board of Education–approved Science Connectors and test blueprints.

7.8.1.3. Intended Test Population

Only eligible students may participate in the administration of the CAA for Science. Any student identified for alternate testing in grades five and eight and in high school (grade ten, eleven, or twelve) takes the CAA for Science. IEP teams "shall determine when a child with a significant cognitive disability shall participate in an alternate assessment aligned with the alternate academic achievement standards."[8]

[8] S. 1177—114th Congress: Every Student Succeeds Act. 2015. Title 1, Part A, Subpart 1, Section 1111(b)(2)(D)(ii)(I)

7.8.2. Evidence Based on Test Content

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

Chapter 4: Test Assembly contains information on which the test forms administered in 2019–2020 were built.

7.8.2.1. Description of the State Standards

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

7.8.2.2. Embedded Performance Task and Item Specifications

Item specifications describe the characteristics of items that are written to measure each content standard. The specifications for science are described in *Chapter 3: Embedded Performance Task and Item Development and Review*.

7.8.2.3. Assessment Blueprints

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

7.8.2.4. Form Assembly Process

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

7.8.3. Evidence Based on Response Processes

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

- systematic observations of test response behavior;
- showing the relationships of items intended to require demonstrations or applications of knowledge and skills to other measures that require similar levels of cognitive complexity in the content (i.e., teacher ratings of student performance); and
- evaluation of the reasoning processes students employ when solving test items (Embretson, 1983; Messick, 1989).

This type of evidence is used to confirm that the CAA for Science is measuring the cognitive skills that are intended as the objects of measurement and are used by students to respond to the items. For example, the survey questions administered after each embedded PT are

typically analyzed as part of the research agenda, with the goal of understanding the CAA for Science embedded PTs.

This subsection describes analyses performed after typical CAA for Science administrations.

7.8.3.1. Analysis of Testing Time

Testing time for each administration can be evaluated for consistency by examining the expected response processes for the items presented to students. The length of time it takes students to complete a test is collected and analyzed to build a profile describing what a typical testing event looks like for each content area and grade. In addition, variability in testing time is investigated to determine whether a student's testing time should be viewed as unusual or irregular. It should be noted that the CAA for Science is an untimed test.

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

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

Testing time analysis was not conducted following the 2019–2020 administration because of small sample sizes.

7.8.4. Evidence Based on Internal Structure

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

7.8.4.1. Differential Item Functioning

DIF analyses are conducted to assess differences in the item performance of groups of students who differ in their demographic characteristics. Refer to section *7.5 Differential Item Functioning Analyses* for a description of the DIF analyses.

7.8.4.2. Overall Reliability Estimates

The description of reliability analyses is provided in section *7.7 Reliability Analyses*. However, because of the COVID-19 pandemic and the governor's order to suspend testing on March 18, 2020, no reliability analyses were conducted for the 2019–2020 CAA for Science administration.

7.8.4.3. Student Groups Reliability Estimates

Because of the COVID-19 pandemic and the governor's order to suspend testing on March 18, 2020, no student groups reliability analyses were conducted for the 2019–2020 CAA for Science administration.

7.8.5. Evidence Based on Relations to Other Variables

Evidence based on *relations to other variables* can be evaluated using the correlation between the CAA for Science assessment results and variables related to students. Two

variables often analyzed related to the students' results include the CAAs for English language arts/literacy and mathematics and the level of student engagement while taking the embedded PTs. However, because of the small number of students who were able to complete the CAA for Science before testing was stopped in March 2020, the analyses with other variables were not conducted after the 2019–2020 administration.

The relationship between test engagement and student's performance on the CAA for Science can provide additional information on the student's testing experience. The minimal levels of engagement for some students could be related to whether students could access the test content while testing or whether the student had the opportunity to learn the content prior to testing. The student's familiarity with the content or tasks would also impact the student's level of engagement. Some students might be more engaged with familiar content, while other students might be more engaged when the content or task is unique (i.e., a novel experience).

References

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

Bhola, D. S., Impara, J. C., & Buckendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice, 22*, 21–29.

Browne, M. W., & Cudek, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J.S. Long (Eds.) *Testing structural equation models* (pp. 136-62). Newbury Park, CA: Sage Publications.

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

California Department of Education. (2018a). *California Alternate Assessment for Science blueprint*. Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx

California Department of Education. (2018b). *California Next Generation Science Standards Core Content Connectors for Alternate Assessments.* Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/documents/ngssaltconnectors.docx

Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: Holt.

Crocker, L. M., Miller, D., & Franks, E. A. (1989). Quantitative methods for assessing the fit between test and curriculum. *Applied Measurement in Education, 2*, 179–94.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.

Dorans, N. J. (2013). ETS contributions to the quantitative assessment of item, test, and score fairness. *ETS Research Report Series*, i–38.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Dorans, N. J., & Schmitt, A. P. (1993). Constructed response and differential item functioning: A pragmatic approach. In R. E. Bennett & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement* (pp. 135–65). Hillsdale, NH: Lawrence Erlbaum Associates, Inc.

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

Embretson (Whitley), S. (1983). Construct validity: construct representation versus nomothetic span. Psychological Bulletin, 93, 179–97.

Feldt, L.S., & Brennan, R. L. (1989). Reliability. In R. L. Linn (Ed). *Educational Measurement* (3rd Edition), (pp 105–46). Phoenix, AZ: Oryx Press.

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

Hanson, B. A., & Béguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common item equating design. *Applied Psychological Measurement, 26*, 3–24.

Holland, P., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Lawrence Erlbaum Associates.

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

Kane, M. (2006). Validation. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 17–64). Washington, DC: American Council on Education and Praeger.

Kim, S. H., & Cohen, A. S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement, 22*, 131–143.

Lord, F. M. & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, MA: Addison-Wesley Publishing Company.

Mantel, N. (1963). Chi-square tests with one degree of freedom: Extensions of the Mantel-Haenszel procedure. *Journal of the American Statistical Association*, *58*, 690–700.

Mantel, N. & Haenszel, W. (1959). Statistical aspects of the analyses of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719–48.

Martone, A., & Sireci, S. G. (2009). Evaluating alignment between curriculum, assessments, and instruction. *Review of Educational Research, 4*, 1332–61.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149–174.

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (3rd ed.). Washington, DC: American Council on Education.

Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*(2): 159–76.

Rothman, R., Slattery, J. B., Vranek, J. L., & Resnick, L. B. (2002). *Benchmarking and alignment of standards and testing* [Technical Report 566]. Washington, DC: Center for the Study of Evaluation.

Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment, 5*, 299–321.

Zwick, R., Thayer, D. T., & Mazzeo, J. (1997). Descriptive and inferential procedures for assessing differential item functioning in polytomous items. *Applied Measurement in Education, 10*(4), 321–44.

Accessibility Information

Alternative Text for Equation 7.1

P value sub dich equals the fraction with the numerator the sum of X sub ij and the denominator N sub I end fraction.

Alternative Text for Equation 7.2

P value sub poly equals the fraction with the numerator X sub ij and the denominator N sub i times Max of X sub I end fraction.

Alternative Text for Equation 7.3

r sub polyreg equals the fraction beta sub hat times S tot divided by the square root of Beta sub hat squared times s sub tot squared plus 1.

Alternative Text for Equation 7.4

Alpha sub MH equals the numerator open parenthesis the sum sub m of R sub rm times W sub fm divided by N sub tm close parenthesis divided by the denominator open parenthesis the sum sub m of R sub fm times W sub rm divided by N sub tm closed parenthesis.

Alternative Text for Equation 7.5

MH D-DIF equals negative 2.35 times the natural logarithm open bracket alpha sub MH close bracket.

Alternative Text for Equation 7.6

SMD equals the fraction with numerator the sum from m equals 1 to M of N sub fm times E sub f of Y from X equals m and denominator the sum from m equals 1 to M of N sub fm end fraction minus the fraction with numerator the sum from m equals 1 to M of N sub fm times E sub r of Y from X equals m and denominator the sum from m equals 1 to M of N sub fm end fraction equals the fraction with the numerator the sum from m equals 1 to M of D sub fm and the denominator m equals1 to M of N suf fm end fraction.

Alternative Text for Equation 7.7

P sub ih of theta sub j equals:

The numerator exp open parenthesis the sum from v equals 1 to h of Da sub i open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis divided by the denominator open parenthesis 1 plus the sum from c equals 1 to n sub I exp open parenthesis the sum from v equals 1 to c of Da sub I open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis close parenthesis, if score h equals 1, 2, …, n sub i.

P sub ih of theta sub j equals:

1 divided by the denominator open parenthesis 1 plus the sum from c equals 1 to n sub I exp open parenthesis the sum from v equals 1 to c of Da sub I open parenthesis theta sub j minus b sub I plus d sub iv close parenthesis close parenthesis close parenthesis, if score h equals 0.

Alternative Text for Equation 7.8

alpha equals the fraction n over n minus one, times one minus the fraction the sum from i equals one to n, of sigma sub i, squared, over sigma sub x, squared.

Alternative Text for Equation 7.9

stratified alpha equals one minus the fraction of the sum sigma sub xj, squared times one minus the alpha sub j over sigma sub x, squared.

Alternative Text for Equation 7.10

SEM equals s sub t times the square root of 1 minus rho of theta hat theta hat prime.

Chapter 8: Surveys

This chapter describes the development and administration of the survey questionnaires for test examiners to complete for each California Alternate Assessment for Science embedded performance task (PT).

8.1. Survey Design and Development

The surveys were designed and developed by members of the ETS validity research team, whose members have extensive experience in designing and developing surveys.

8.1.1. Survey on the Test Administration

Student survey responses, which were provided by the test examiner, were collected from local educational agencies via the California Assessment of Student Performance and Progress test delivery system (TDS) for every embedded PT administered to every student. After the embedded PT was administered to the student, the test examiner was presented with two surveys about the student, with the instruction to only respond to one of the surveys based on whether the student had been responsive during the testing session. The purpose of the survey was to collect basic information about students' experiences with the assessment process.

The survey was included in the last section of the embedded PT delivered through the TDS. However, because of the impact of the novel coronavirus disease 2019 pandemic, no analyses of the survey responses were conducted for the 2019–2020 administration.

8.1.1.1. Survey for Responsive Students

The survey for responsive students consisted of the following three questions:

1. How engaged was the student with this performance task?
    - Fully engaged
    - Moderately engaged
    - Minimally engaged
2. Did you individualize any aspect of Orienting Activity #1 and the first five test questions, where permitted? If yes, and you used specific materials, briefly describe.
3. Did you individualize any aspect of Orienting Activity #2 and the last five test questions, where permitted? If yes, and you used specific materials, briefly describe.

8.1.1.2. Survey for Nonresponsive Students

The survey for the nonresponsive students consisted of the following four questions:

1. Which of the following statements best explains why your student did not provide any response?
    - No established mode of communication
    - No observable engagement with the performance task
    - Test questions seem to be too complex
    - Scientific concepts seem to be too complex
    - Other

2. During classroom instruction, what mode or modes of communication does your student use? Select all that apply.

    - Mouse, touchscreen, computer keyboard, or any combination of these
    - Verbal response
    - Gestures or pointing
    - Written response
    - Augmentative and alternative communication device
    - Eye gaze
    - Nonresponsive
    - Other

3. What method(s) did you use to elicit a response from your student?

4. How might this embedded performance task be changed so your student could be successful in showing what they know and can do?

Chapter 9: Quality Control Procedures

The California Department of Education (CDE) and ETS implemented rigorous quality control procedures throughout the test development, administration, scoring, analyses, and reporting processes associated with the California Alternate Assessment (CAA) for Science. As part of this effort, ETS California Assessment of Student Performance and Progress (CAASPP) program staff worked with the ETS Office of Professional Standards Compliance, which publishes and maintains the *ETS Standards for Quality and Fairness* (ETS, 2014). These *Standards* support the goals of delivering technically sound, fair, and useful products and services; and assisting the public and auditors evaluating those products and services. Quality control procedures are outlined in this chapter.

9.1. Quality Control of Embedded Performance Task Development

ETS' goal is to provide the best standards-based embedded performance tasks (PTs) for the CAA for Science. Embedded PTs developed for the CAA for Science underwent an extensive embedded PT review process. The item writers hired to develop CAA items were trained in CAASPP and ETS policies on quality control of item content, sensitivity, and bias guidelines, as well as on guidelines for accessibility, to ensure that the items allow the widest possible range of students to demonstrate their content knowledge.

Once a written item was accepted for authoring—that is, once it was entered into ETS' item bank and formatted for use in an assessment—ETS employed a series of internal and external reviews. These reviews used established criteria and specifications to judge the quality of items and to ensure that each item measured what it was intended to measure. These reviews also examined the overall quality of the test items before they were presented to the CDE and item reviewers. Finally, a group of California educators reviewed the items for accessibility, bias and sensitivity, and content prior to their administration to students. The details on quality control of item development are described in section *3.2 ETS Item Review Process*.

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. The blueprints for the CAA for Science were approved in January 2018 and test forms were assembled following the approved blueprints (CDE, 2018).

9.2.1. Quality Control of Test Form Development

ETS conducted multiple levels of quality assurance (QA) checks on each constructed test form to ensure it met defined statistical criteria. For the 2019–2020 CAA for Science, both ETS Assessment and Learning Technology Research & Development (ALTRD) and psychometric staff reviewed and signed off on the accuracy of forms before the test forms were put into production. Detailed information related to test assembly can be found in *Chapter 4: Test Assembly*.

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

- all correct answers were correct;
- items were scored correctly in the item bank;
- all embedded PTs aligned with the standard;
- all content in the embedded PT was correct;
- distractors were plausible;
- multiple-choice item options were parallel in structure;
- language was grade-level appropriate;
- no more than three multiple-choice items in a row had the same key;
- all graphics were correct (copyright, spelling, relevance, etc.);
- there were no errors in spelling or grammar; and
- embedded PTs adhered to the approved style guide.

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

9.2.2. Quality Control of Test Assignment

Test assignment for the CAASPP, including the CAA for Science, is controlled by the Test Operations Management System (TOMS) using student information received from the California Longitudinal Pupil Achievement Data System (CALPADS) (CDE, 2020). The two systems are kept in sync during the testing window.

Students at eligible grade levels were assigned to the Smarter Balanced assessments (in grades three through eight and grade eleven) and the California Science Test (CAST) (grades five and eight and high school) by default. For students eligible for the CAA for Science—that is, grades five and eight and high school (grade ten, eleven, or twelve)—local educational agencies (LEAs) logged on to TOMS and assigned students to take the alternate assessment, which automatically unassigned those students from taking the CAST.

The quality of test assignment for the CAA for Science was monitored and controlled through several strategies. TOMS enforced preconditions for eligibility for the CAAs by permitting assignment only for students with an Individuals with Disabilities Education Act[9] indicator of "Yes" in TOMS. This indicator is set to "Yes" when the CALPADS *Education Program* field (field 3.13) is equal to 144 (Special Education) and the primary disability code (field 3.21) is not set to blank.

Additionally, TOMS prevented the prohibited "mixing and matching" of assessments. For example, a student assigned to take an alternate assessment for any content area will automatically be prevented from assignment to a general assessment for another content area.

[9] 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 received training on these topics and were provided with tools to report security incidents and resolve testing discrepancies for specific testing sessions.

Several strategies are utilized to monitor and control the quality of test administration for the CAA for Science as well as all assessments administered as part of the CAASPP System. A fully staffed support center, the California Technical Assistance Center (CalTAC), supports all LEAs in the administration of all CAASPP assessments. CalTAC is guided by a core group of LEA outreach and advocacy staff who manage communications to LEAs, regional and web-based trainings, and a website, https://www.caaspp.org/, that houses a full range of manuals, videos, and other instructional and support materials. In addition to providing guidance and answering questions, CalTAC regularly conducts outreach campaigns on particular administration topics to ensure all LEAs understand correct test administration procedures.

The ETS Office of Testing Integrity (OTI) reinforces the quality control procedures for test administration, providing QA services for all testing programs managed by ETS. The OTI's detailed quality control procedures are described in subsection *5.7.1. ETS' Office of Testing Integrity*.

9.2.4. Quality Control of Machine-Scoring Procedures

To ensure valid item-level scoring for the CAA for Science, quality control procedures were employed by Cambium Assessment, Inc. (CAI), the CAASPP subcontractor responsible for providing the CAASPP test delivery system (TDS) and scoring machine-scorable items. A final comparison of the test map to each online form as configured in the UAT environment ensured that no changes to the form were introduced prior to operational deployment.

A real-time, quality-monitoring component was built into the TDS. After a test was administered to a student, the TDS passed the resulting data to the QA system. QA conducted a series of data integrity checks, ensuring, for example, that the record for each test contained information for each item, keys for multiple-choice items, score points in each item, and the total number of operational items. In addition, QA also checked to ensure that the test record contained no data from items that had been invalidated.

Data passed directly from the quality monitoring system to the database of record, which served as the repository for all test information, and from which all test information for reporting was pulled and transmitted to ETS in a predetermined results format.

9.3. Quality Control of Test Materials

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

9.3.1. Test Administration Manuals

ETS staff consulted with internal subject matter experts and conducted validation checks to verify that test directions and administration manuals accurately matched the test materials and testing processes. Copy editors and content editors reviewed each document for spelling, grammar, accuracy, and adherence to CDE style. Each document was required to be approved by the CDE before it could be published to the CAASPP website at

https://www.caaspp.org/. Only nonsecure documents were posted to this website. Secure materials, such as the *CAA for Science Embedded Performance Task Directions for Administration*, were made available to designated LEA staff through TOMS, which required a secure logon.

The manuals used in the administration of the CAA are listed in subsection *5.4.4 Instructions for Test Examiners and Staff Involved in CAA for Science Administration*.

9.3.2. Processing Test Materials

The following information was entered into the TDS by test examiners and transmitted from CAI to ETS each day:

- Student's first name
- Statewide Student Identifier
- Any individualized scripts and materials used
- Responses for each item
- Results of the student survey
- Results of the responsive student survey or the nonresponsive student survey

The CAI 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 were completed, approved, and checked well in advance of the receipt of student response data. These specifications contained detailed scoring procedures, as well as the procedures for determining whether a student attempted a test and whether that student's response data should be included in the statistical analyses and calculations for computing summary data.

9.4.2. Development of Scoring Procedures

ETS' Enterprise Score Key Management (eSKM) system utilized scoring procedures specified by psychometricians and provides scoring services. ETS psychometricians carried out a series of quality control checks after scoring to ensure the accuracy of each score.

9.4.2.1. Enterprise Score Key Management System Processing

Prior to the test administration, ETS ALTRD staff reviewed and verified the keys for all items. Then, these keys were provided to CAI for its machine-scoring implementation. After CAI finished machine-scoring, those scores and responses were delivered to ETS. CAI quality control of the machine-scoring procedure is described in subsection *9.2.3 Quality Control of Test Administration*.

ETS' Centralized Repository Distribution System and Enterprise Service Bus departments collected and parsed .xml files that contained student response data from CAI. ETS' eSKM system collected and calculated individual students' overall scores (total raw scores) and generated student scores in the approved statistical extract format. These data extracts were sent to ETS' Data Quality Services for data validation. Following successful validation, the student response statistical extracts were made available to the psychometric team.

ETS developed two parallel scoring systems to produce and verify student scores:

4. The eSKM scoring system received an individual student's item scores and item responses from CAI and calculated individual student scores for ETS' reporting systems.
5. The ETS Psychometric Analysis & Research team computed individual student scores based on item scores delivered by CAI.

The scores from the two sources were then compared for internal quality control. Any differences in the scores were discussed and resolved. All scores complied with the ETS scoring specifications and passed the parallel scoring process to ensure the quality and accuracy of scoring and to support the transfer of scores into TOMS, the database of the student records scoring system.

9.4.2.2. Psychometric Processing

No psychometric analyses were conducted on the data from the 2019–2020 administration. In a typical year, when the psychometric analyses are conducted at ETS, the psychometric analyses undergo comprehensive quality checks by a team of psychometricians and data analysts. Detailed checklists are developed by members of the team for each statistical procedure performed on the CAA for Science.

The ETS psychometrics team reviews the data files before conducting the statistical analyses to ensure the quality of the data. 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 ALTRD staff for review. ALTRD comments are reviewed by the psychometricians before items are approved for inclusion in additional analyses and before the data review meetings with the CDE.

During the calibration process, checks are made to ascertain that the input files are established accurately. Checks are also made on the number of items, number of examinees with valid scores, item response theory (IRT) item difficulty estimates, and standard errors for the item difficulty estimates. Two psychometricians conduct parallel calibration processing and compare the results to check for any inconsistencies. Psychometricians also perform detailed reviews of relevant statistics to determine whether the chosen IRT model fit the data. In addition, the results of the calibration procedures are reviewed by a psychometric manager.

9.5. Quality Control of Reporting

A data file summarizing the results of the 2019–2020 administration for the CAA for Science was provided to the CDE. This file included the percent correct and preliminary indicator category for students who completed the CAA for Science. To ensure the quality of the data file, two members of the ETS psychometrics team individually reviewed the files and worked with ETS Information Technology to resolve any discrepancies before the files were posted to the CDE.

9.6. End-to-End Testing for Operational Administration

ETS conducted end-to-end testing prior to the start of the test administration. The purpose of this testing was to verify that all systems, processes, and resources were ready for the operational administration. ETS employed a number of approaches to verify ongoing systems performance, including monitoring of system availability and online system usage. Time was allotted for UAT to confirm that the systems met requirements and to make identified corrections before final deployment. To accomplish system acceptance and sign off, ETS deployed systems to a staging area, which mirrored the final production environment, for operational testing and UAT. Final approval by the CDE triggered the final deployment of the system.

References

California Department of Education. (2018). *California Alternate Assessment for Science blueprint*. Sacramento, CA: California Department of Education. https://www.cde.ca.gov/ta/tg/ca/documents/caascienceblueprint.docx

California Department of Education. (2020). *CAASPP and ELPAC Test Operations Management System user guide*. Sacramento, CA: California Department of Education. https://bit.ly/2QKKJaV

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

Chapter 10: Continuous and Systematic Improvements

This chapter discusses the various procedures used to gather information to improve the California Alternate Assessment (CAA) for Science as well as strategies to implement possible improvements.

10.1. Item Development

In 2019, ETS led a text complexity meeting with California Department of Education (CDE) staff to discuss and evaluate tools and processes for determining appropriate levels of text complexity in stimuli and items. The structure of the stimuli and item text on the CAAs presents challenges to the use of some common text complexity evaluation tools. A holistic array of tools and rubrics were proposed for use to provide a more accurate measure of the appropriateness of the language of stimuli and items. These measures are now used as part of the item development process for the CAAs.

10.2. Test Delivery

One improvement being implemented is a routing question to be provided before the survey at the end of the test. Therefore, instead of test examiners being presented with two surveys and choosing the correct one to respond to on the basis of their student's responsiveness, the routing question will lead the test examiner to the correct survey and only the correct survey.

Information pages for the orienting activities are also being added. Previously, only orienting activities with videos had information pages within the test delivery system (TDS), but going forward, each orienting activity will have a corresponding information page to ensure test examiners administer the orienting activity.

The *Mark as No Response* option, available on every item within the TDS, was added to provide information regarding the student testing experience. The *Mark as No Response* option should be used when the item is presented to the student and the student does not provide a response despite the test examiner's best efforts to elicit a response. Previously, a test examiner would have just skipped ahead in the test; this option allows for greater detail in the data gathered from the test.

10.2.1. Stakeholder Feedback

The California Assessment of Student Performance and Progress (CAASPP) program solicits feedback annually from various stakeholder groups, including local educational agency (LEA) CAASPP coordinators, CAASPP test site coordinators, test administrators, and test examiners, through the CAASPP Post-Test Survey. In 2018–2019, feedback was collected via a post-test survey of more than 8,000 California educators and a focus group study with over 40 participants. (There were not enough responses received about the CAA for Science from the 2019–2020 survey to provide analysis.)

10.2.2. Commitment to Preparation and Training Resources

When asked about preparation and training resources for the CAAs, 2018–2019 survey respondents found them to be helpful overall. When asked if the online, self-guided Test Examiner Tutorial provided all the training necessary for CAA test examiners, 87 percent of

all respondents agreed it did. More than 80 percent of respondents found the tutorial to be helpful in preparing them to administer the CAAs.

The Test Examiner Checklists were found to be helpful by 65 percent of respondents; the CAA For Science Webcast and the Practice Test Scoring Guides were the least-accessed materials.

To improve the dissemination of information to all educators involved in the CAAs, ETS plans to communicate more information to LEA CAASPP coordinators and CAA test examiners at the same time. The survey results support this strategy, with 48 percent of LEA CAASPP coordinators delegating CAA testing activities to special education staff and 55 percent agreeing that ETS should send CAA communications to all relevant audiences.

10.3. Psychometric Analyses

ETS analyzes CAA for Science items, including the following types of analyses:

- Item difficulty
- Item discrimination
- Item score distributions
- Differential item functioning

The CDE requested that the item difficulty flagging criteria be based on the number of response choices. Therefore, for future test administrations, ETS will flag difficult multiple-choice items based on the number of response options for the item:

- Items with only two response options will be flagged when the item's *p*-value is below 0.50.
- Items with three response options will be flagged when the item's *p*-value is below 0.30.
- Items with four response options will be flagged when the item's *p*-value is below 0.20.

The polytomous items will continue to be flagged when the proportion of correct responses is less than 0.30.

10.4. Research-based Operational Work

A feature of the CAA for Science embedded performance tasks (PTs) is that the test examiners have the option to individualize certain elements of the assessment, although not all embedded PTs allow for individualizations. Potential individualizations are designed so that the premise of the item and the scientific principles tested would remain the same. Individualization options in embedded PTs often involve the use of objects to make certain science concepts easier to understand for some students. Therefore, there were concerns about the potential impact of giving test examiners the flexibility to choose materials to conduct activities associated with the embedded PTs.

ETS evaluated the impact of the individualizations after the 2017–2018 second-year pilot and after the 2018–2019 field test administrations. In general, individualizations and material choice do not explain a significant proportion of the variance of the students' California Next Generation Science Standards Core Content Connector (Science

Connector) scores. Student engagement and student disability explained significant proportions of the Science Connector scores.

When interpreting the results of the material choice analyses, caution should be taken because of the small percentage of students who received an individualization (e.g., individualized script) or who received individualized materials. Because of the low rates of students receiving an individualization, there is low statistical power to detect possible effects of the choice to use individualized materials if an effect exists. Additionally, the test examiner chose to use individualizations or individualized materials to make the Science Connector orienting activity more accessible to the student and was based on the needs of the student. Therefore, the results of these analyses are nested within student disability and the needs of the student.

The ETS psychometricians will continue to monitor the number of students receiving an individualization and, for future administrations, will evaluate the impact of the individualizations on student's scores.

10.5. Accessibility

Like all CAASPP assessments, the CAAs are administered using the TDS created by Cambium Assessment, Inc. for the Smarter Balanced assessments. As such, implementation of new online universal tools, designated supports, and accommodations are aligned with the TDS.