2020 No. 085

California Assessment of Student Performance and Progress (CAASPP): 2020 Independent Evaluation Report (Volume 1)

Prepared for: California Department of Education
Assessment Development
and Administration Division
1430 N Street, Suite 4401
Sacramento, CA 95814-5901

Prepared under: CN180100.1

Authors: Michele M. Hardoin
Rebecca Norman Dvorak
Emily Dickinson
Justin Paulsen
Monica Gribben
Rivka Revivo

Editors: Sheila Schultz
Laurie Wise
Christa Watters

Date: October 30, 2020

Headquarters: 66 Canal Center Plaza, Suite 700, Alexandria, VA 22314 | Phone: 703.549.3611
https://www.humrro.org

This page is intentionally blank.

CAASPP 2020 Independent Evaluation Report

Table of Contents

List of Tables

List of Figures

This page is intentionally blank.

Executive Summary

Pursuant to California *Education Code* (*EC*) Section 60649, the Human Resources Research Organization (HumRRO) is continuing its independent evaluation of the California Assessment of Student Performance and Progress (CAASPP) System. In addition to assessments, the CAASPP System includes resources to help students, parents/guardians, teachers, and administrators understand students' progress toward meeting grade-level standards. This annual report covers the activities HumRRO conducted during the 2019–2020 academic year for each of the following studies:

- Instruction and Student Learning Case Study, year two
- California Science Test (CAST) Alignment Study
- California Alternate Assessment (CAA) for Science Alignment Study

For each study, this executive summary provides an overview, a summary of the major findings, and recommendations for improvement to the studied CAASPP components. Detailed descriptions and findings of each study are presented in chapters 2 through 4, and a more in-depth presentation of conclusions and recommendations is presented in chapter 5.

Instruction and Student Learning Case Study

Overview

The primary goal of the two-year case study is to elicit concrete examples of how and why specific Smarter Balanced English language arts (ELA) and mathematics components are used and the perceived benefits and challenges of using them. The three components are the summative assessments; interim assessments (IAs), which include Interim Assessment Blocks (IABs) and longer Interim Comprehensive Assessments (ICAs); and the Digital Library (DL), which includes formative assessments and instructional resources. [1] For year two, HumRRO collaborated with six LEAs, including one direct-funded charter, and a subset of their schools (15 schools in all) to study their use of the Smarter Balanced components.

Summary of Findings

The following high-level summaries describe how educators across the small sample of LEAs and schools in the study used the Smarter Balanced components during the 2019–2020 academic year:

[1] During the period of this study, the DL was the system available to educators. The DL was retired in May 2020 and replaced by a new system (Tools for Teachers) in June 2020.

- **Summative Assessments**. Most school staff participating in the study reviewed summative assessment data from the prior year, often as a school-wide or grade-level team. Almost all school leaders and teachers at the elementary and middle schools (a) reviewed achievement level results by grade and (b) compared performance across similar districts and schools. Some schools also reviewed more detailed results (e.g., claims) and used those results to help identify annual achievement goals or influence instructional foci or the selection of IABs administered during 2019–2020.

- **Interim Assessments.** All schools in the study used IAs in both ELA and mathematics, except for one elementary school. Some LEAs mandated IA use, and some of these LEAs also specified which IAs were to be administered per subject and grade level. Many teachers cited benefits of IAs for monitoring student progress and informing instructional decisions, beyond their usefulness for preparing students for the summative assessments. The most positive perceptions about IABs were from teachers who had input into decisions about giving IABs, which allowed better alignment of assessments with their curriculum.

- **Digital Library.** The study schools reported extremely limited use of the resources of the Digital Library, though most teachers were aware of the resources and had logged in at least once. Many teachers noted time constraints, difficulty finding useful resources, difficulty navigating through the system, and availability of sufficient materials through their curriculum or other familiar sources as reasons for not using the DL.

Summary of Best Practices

Across the studied LEAs and schools, HumRRO identified the following sample of best practices used by participating LEAs for successful implementation of the Smarter Balanced components. Each “best practice” is an approach that (a) aligned well with the intended purpose of and guidance for implementing components within the CAASPP System and (b) resulted in educators having a positive experience using the CAASPP System to inform their teaching.

- Use summative assessment data to monitor school-level performance and, in combination with other data, to identify school-wide goals.

- Use IAs as a teaching tool. For example, review commonly missed items as a full class, small group, or partner exercise.

- Use IA data to identify gaps in student understanding and determine content that should be retaught to the full class or select groups of students.

- Provide support and training at the school and local educational agency (LEA) levels for using CAASPP resources.

- Provide leadership guidance and encouragement for using CAASPP components while allowing teachers flexibility regarding which IAs and DL resources to incorporate into their classrooms.

- Facilitate school-wide data discussions to ensure teachers know how to access and interpret summative assessment results, and how these data can inform instructional practices.

- Provide time and resources to support collaboration among grade-level and/or content-area professional learning communities (PLCs) to plan instruction and use interim and formative assessments effectively.

Recommendations

HumRRO reviewed the full scope of study findings to develop suggestions for the CDE to consider as part of its continuous improvement of the CAASPP System. Based on data from year two—from a small number of teachers within a small number of schools in a small number of LEAs—we offer the following recommendations to the CDE. Most of these are already being addressed by significant enhancements the CDE will implement during the 2020–2021 school year. Expanded recommendations and descriptions of enhancements being implemented are described in chapter 5.

Recommendation 1: Continue providing training opportunities and updated online resources for LEA- and school-level staff.

Recommendation 2: Work with the Smarter Balanced Assessment Consortium to provide an expanded pool of ELA and mathematics IAs, particularly Focused IABs (FIABs), and develop multiple versions of existing IAs.

Recommendation 3: Use the CAASPP website to address the issues of version control and changing CAASPP component guidance to ensure educators are aware of new releases and use current resources.

Recommendation 4: Consider adding reporting elements and resources directed toward students at the upper grade levels, providing them with information and tools to enhance their own learning.

Recommendation 5: Continue efforts to increase usability of online platforms.

Recommendation 6: Seek ways to improve online access to high quality, free, Common Core State Standards (CCSS)-aligned formative assessment resources for school-level staff.

California Science Test Alignment Study

Overview

The CAST is a computer-based assessment administered to students in grades five, eight, and once in high school (i.e., grades 10, 11, or 12). The CAST is designed such that its content at each grade level will rotate annually across a three-year span, each year sampling different content from the California Next Generation Science Standards (CA NGSS) to allow CAST to address the full breadth of the standards. Within the CA NGSS, performance expectations (PEs) are assessable statements of what students should know and be able to do. Three major components, also referred to as dimensions, are combined to operationalize the PEs: Disciplinary Core Ideas (DCIs), Crosscutting Concepts (CCCs), and Science and Engineering Practices (SEPs).

For the CAST alignment study, HumRRO conducted two major activities. First, we evaluated the degree to which the CAST test design and development documentation met relevant testing standards in the *Standards for Educational and Psychological Testing* (AERA, et al., 2014). HumRRO then collected evidence of whether the CAST produces test **forms** that effectively measure the content and cognitive rigor reflected in the targeted content domain (CA NGSS) and the test blueprints. Alignment studies are required as part of the federal assessment peer review process, provide validity evidence that the assessment is measuring the intended content, and inform future assessment item development.

Summary of Findings

This section provides a high-level summary of the findings from the two major study activities HumRRO conducted to evaluate the alignment between the CAST and the CA NGSS, documentation review and item ratings by content experts:

- The test design and test blueprints for the CAST support the conclusion that the testing contractor adhered to testing standards relevant to test-to-standards alignment.
- Review of operational test forms from the 2018–2019 administration support the claim that the CAST design produces aligned test forms.
- The PEs assessed via the 2018–2019 item pool are sufficient to support the claim that the CAST is on track to address the full breadth of the CA NGSS after two additional operational administrations.
- The number of items linked to each content domain, SEP, and CCC align with the guidelines presented in the CAST blueprints. In only a small number of instances did the number of items rated as aligned to a particular dimension fall slightly outside of the ranges specified in the blueprint.

- Most of the CAST items, across the grade levels, are multidimensional (i.e., measure a PE by integrating a DCI, CCC, and/or SEP). Across the grade levels, the majority of items were rated as multidimensional, and more than half of items on any test form were rated as integrating all three dimensions.

- CAST forms across the grade levels reflect reasonable balance across the disciplinary areas used for scoring and reporting purposes (Earth and Space Sciences, Life Sciences, and Physical Sciences), as well as across the CA NGSS SEPs and CCCs.

- CAST items vary in cognitive complexity, with slightly more than 10 percent at Level 1 Depth of Knowledge (DOK) and more than 10 percent at Level 3 DOK.

- For all three grades, the distribution of item difficulties generally lines up with the distribution of student ability levels.

Recommendations

The study results were generally very positive and do not indicate that any major changes in test development or forms construction processes and procedures are needed. We offer one recommendation for improving the CAST blueprints.

> **Recommendation 1**: Add recommended cognitive complexity distributions to the CAST blueprints, along with a rationale for the targets set for each level.

California Alternate Assessment for Science Alignment Study

Overview

The CAA for Science is administered to eligible students in grades five, eight, and once in high school (i.e., grades 10, 11, or 12). Individualized education program (IEP) teams "shall determine when a child with the most significant cognitive disability shall participate in an alternate assessment aligned with the alternate academic achievement standards." (Title 1, Part A, Subpart 1, Sec. 1111(b)(2)(D)(ii)(I)—Every Student Succeeds Act, 2015). The CAA for Science is designed to measure performance on the Science Connectors, which are derived from the performance expectations (PEs) of the CA NGSS. The CAA for Science is not a single end-of-year summative test but instead is designed to be administered as four separate sessions following instruction throughout the school year. Each session consists of one performance task (PT), and each PT addresses one science domain (i.e., Earth and Space Sciences, Life Sciences, and Physical Sciences). The students' performance on the three operational PTs are aggregated to generate an overall science score at the conclusion of the school year. The fourth PT is for field test purposes.

The CAA for Science was field tested in the 2018–2019 school year and was to be administered operationally for the first time in 2019–2020. However, CDE received a waiver for accountability testing from the Federal Government in 2019–2020 due to

COVID-19 school closures. As a result, the first operational administration of the CAA for Science was delayed until the 2020–2021 school year.

For the CAA for Science alignment study, HumRRO conducted two major activities. First, we evaluated the degree to which the CAA for Science test design and development documentation met relevant testing standards in the *Standards for Educational and Psychological Testing* (AERA, et al., 2014). HumRRO then collected evidence of whether the CAA for Science produces test **forms** that effectively measure the content and cognitive rigor reflected in the targeted content domain (Science Connector) and the test blueprints.

Summary of Findings

This section provides a high-level summary of the findings from the two major study activities HumRRO conducted, documentation review and item ratings by content experts, to evaluate the alignment between the CAA for Science and the Science Connectors derived from the CA NGSS.

- The test design and test blueprint for the CAA for Science support the conclusion that the testing contractor adhered to testing standards relevant to test-to-standards alignment. Review of items that were ready for operational use in 2019–2020 supports the claim that the CAA for Science design produces aligned test forms.

- All performance tasks in each of the three content domains were linked to at least two Science Connectors, as outlined in the test blueprint.

- For all three CAA for Science tests (grade five, grade eight, and high school), all items were judged as being aligned to a Science Connector. Similarly, all PTs at all three grade levels measured multiple Science Connectors, Essential Understandings (EUs), and Focal Knowledge, Skills, and Abilities (FKSAs). Regardless of the version administered, every student was tested via a form that fully met the Link to Standards and Range Adequacy criteria.

- For all three grade-level CAA for Science tests, items were rated at each of the three levels of cognitive complexity. The number of items rated at each level of cognitive complexity fell within appropriate ranges for the item pools of all three grade-level tests.

- For all grades, test form versions generally included appropriate numbers of items from each cognitive complexity level.

Recommendations

The study results were generally very positive and do not indicate that any major changes in test development or form construction processes and procedures are needed.

Summary and Next Steps

With the conclusion of the 2019–2020 academic year, the remaining activity for HumRRO's CAASPP independent evaluation is to prepare (a) a stand-alone Instruction and Student Learning Case Study report for year two and (b) a final comprehensive report for the 2018–2020 evaluation. HumRRO has been honored to be the independent evaluator for CDE's assessment programs since 1999, contributing our objective and high-quality research efforts to support the continuous improvement of first the California High School Exit Examination and now the CAASPP System.

This page is intentionally blank.

Chapter 1: Introduction

Background

The California Assessment of Student Performance and Progress (CAASPP) System has been the statewide student assessment program since 2014. It is intended to assist teachers, administrators, students, and parents by promoting high-quality teaching and learning using a variety of assessment approaches and item types. The Smarter Balanced English language arts/literacy (ELA) and mathematics tests monitor progress in implementing effective instruction aligned with the Common Core State Standards (CCSS) for ELA and mathematics; the California Alternate Assessments in ELA and mathematics have been operational since 2016. The California Science Test (CAST), which became operational in spring 2019, and the California Alternate Assessment in Science (CAA Science), which will be operational during the 2020–2021 school year, are aligned to the California Next Generation Science Standards (CA NGSS). The CAASPP System also includes an optional Spanish reading language arts test, the California Spanish Assessment (CSA), which became operational in 2019. These assessments aim to shift the focus away from accountability toward a comprehensive plan for promoting teaching and learning for all students, including students with disabilities (SWDs) and English learners (ELs). The CAASPP System includes sophisticated online tools for reporting assessment results and represents a substantial financial investment by the state as well as a significant investment of educator and student time.

California *Education Code* (*EC*) Section 60649(a) requires the independent evaluation of the CAASPP System, stating that "evaluation activities may include a variety of internal and external studies such as validity studies, alignment studies, and studies evaluating test fairness, testing accommodations, testing policies, and reporting procedures, and consequential validity studies specific to pupil populations such as English learners (ELs) and pupils with disabilities." The law requires development of a plan to conduct independent evaluation activities, and it prohibits duplication of studies conducted as part of a federal peer review process or by California Department of Education (CDE) assessment contractors.

The Human Resources Research Organization (HumRRO) served as the first CAASPP System evaluator from 2015 to 2018. Copies of our annual and comprehensive final reports are available on the CDE Web page (https://www.cde.ca.gov/ta/tg/ca/caaspprptstudies.asp).

The CDE awarded the contract for the 2018–2020 independent evaluation of the CAASPP System to HumRRO in July 2018. The current contract calls for annual evaluation reports that summarize all work completed during the previous year, stand-alone reports for individual research studies, and a comprehensive final report. Within a few months of the award, HumRRO submitted to the CDE the first required annual evaluation report (Hardoin et al., 2018). That report's core contents included the 2018–2020 Evaluation Plan, which described the design of three research studies approved

by the CDE and scheduled within the contract period. The present report is the third annual report and describes results from the studies concluded during 2019–2020. A Comprehensive Final Evaluation Report 2018–2020 will be delivered in 2020 and will summarize evaluation findings and recommendations from each of the three annual reports.

During the 2019–2020 academic year, the Coronavirus disease (COVID-19) outbreak had a significant impact on the CAASPP System and the delivery of instruction at all grade levels across the state. All CAASPP testing was suspended on March 20, 2020. Local educational agencies (LEAs), schools, and teachers who had been implementing the various components of the expanded CAASPP System faced widespread school closures and conversion to full- or part-time distance learning, which impacted the Instruction and Student Learning Case Study but did not affect the alignment studies.

An ongoing evaluation is important to support the goal of continuous improvement to help California achieve the intended return on its investment in the CAASPP System. The evaluation can provide evidence to demonstrate the validity of intended interpretations of test scores used as measures of student learning relative to targeted content standards. It also can offer recommendations for potentially improving alignment between what an assessment is intended to measure and what it actually measures. The evaluation can also provide insight into how CAASPP results are used to improve instruction at the student, classroom, school, local educational agency (LEA), and statewide levels.

2018–2020 Evaluation Plan Goals and Timeline

As context for this year's report on evaluation activities, table 1.1 presents an overview of the goals and schedule for each research study included in the 2018–2020 Evaluation Plan. HumRRO developed the plan with guidance from the CDE and input from the CAASPP Technical Advisory Group (TAG). Each research study was designed to provide information about how well specific elements of the CAASPP System as delivered meet the intended goals of the program expressed in the CAASPP System theory of action. The plan in its entirety is available in the *2018 CAASPP Evaluation Report* (Hardoin, M. M., et al., 2018).

Table 1.1 Overall Goals and Schedule for Each 2018–2020 Evaluation Study

Study Title	Goals and Schedule
Instruction and Student Learning Case Study	• Collaborate with and gather extensive qualitative data (case studies) from a small sample of schools and LEAs, purposefully selected based on their use of CAASPP components and resources. The small sample will aim to broadly represent the diversity of the state with respect to geographic location, academic achievement, and size (student enrollment), as well as student population characteristics (i.e., socioeconomic disadvantage and EL status).

Table 1.1 (cont.)

Study Title	Goals and Schedule
Case Study *(cont.)*	• Investigate the context and various approaches used by the small sample of schools and LEAs to implement and integrate the CAASPP System components to inform instruction and improve student learning. Case Study reports will each describe in detail one school year's findings of the studied LEAs' and schools' use of CAASPP components and the impact of each component on instruction and student learning. The report will document in detail the local context for each case study. • Conduct year one data collection activities in 2018–2019 with initial set of LEAs and schools. • Complete year one data analysis in 2019 and develop stand-alone year one report. • Conduct year two data collection activities in 2019–2020 with second set of LEAs and schools. • Complete year two data analysis in 2020 and develop stand-alone year two report.
CAST Alignment Study	• Evaluate the degree of alignment between the CAST test items and test forms with the CA NGSS. • CAST Alignment Study Report should guide future item development and provide validity evidence suitable for submission for federal peer review under the Every Student Succeeds Act (ESSA). • Conduct data collection activities in 2018–2019. • Complete data analysis in 2020 and develop stand-alone report.
CAA for Science Alignment Study	• Evaluate the degree of alignment between the test items and test forms of the CAA for Science with the Science Connectors and Focal Knowledge, Skills, and Abilities (FKSAs) derived from the CA NGSS. • CAA for Science Alignment Study Report should guide future item development and provide validity evidence suitable for submission for federal peer review under ESSA. • Conduct data collection activities in 2019–2020. • Complete data analysis in 2020 and develop stand-alone report.

Implementing the Evaluation Plan in 2019–2020

A summary list of key Evaluation Plan activities conducted during 2019–2020 is presented in table 1.2, along with the status of the work as of June 30, 2020.

Table 1.2 Schedule and Status of Evaluation Activities for 2019–2020

Activity	Time Frame	Status
Management Meetings with CDE Staff: Biweekly calls to discuss progress, plans, and issues.	July 2019–June 2020	Completed
State Board of Education (SBE) Meetings: Meet with SBE staff and provide presentations at Board meetings.	As requested, up to two times annually	Not Scheduled
TAG Meetings: Meet with and provide presentations to the CAASPP TAG, including detailed designs, review of progress on studies, preliminary findings from studies, and Evaluation Plan updates.	Three times annually, July 2019–June 2020	Completed
CAASPP Contractor Annual Planning Meeting: Attend meeting to learn of planned updates to the system, concerns, processes, scope, and schedule.	Semiannually, July 2019–June 2020	Completed
Develop and deliver the stand-alone report for the CAST Alignment Study.	July 2019–June 2020	Completed
Develop and deliver the stand-alone report for year one of the Case Study.	July–December 2020	Completed
Conduct year two of the Case Study.	July 2019–June 2020	Completed
Observe two CAASPP Smarter Balanced educator training sessions.	July 2019–June 2020	Completed
Conduct the CAA for Science Alignment Study; develop and deliver a stand-alone report.	July 2019–June 2020	Completed
Develop and deliver the 2019 annual report.	July–December 2020	Completed
Maintain comprehensive plan and schedule for project activities and deliverables.	July 2019–June 2020	Completed
Submit monthly written progress reports to describe evaluation progress, plans, and issues.	July 2019–June 2020	Completed

Background Research on Updated CAASPP System

The CAASPP system has continued to evolve during the course of this independent evaluation. Following are important changes and additions implemented by the CDE during 2019–2020 that affected HumRRO's studies:

- 2019–2020 CAA for Science administration, starting in September 2019
- Addition of 73 more ELA and mathematics Smarter Balanced Digital Library (DL) Connections Playlists, providing links to DL resources on the basis of students' Interim Assessment Block (IAB) performance
- Launch of 40 new Smarter Balanced Focused IABs (FIABs) for ELA and math, and corresponding DL Connections Playlists
- Transition from CAASPP Test Operations Management System (TOMS) to MyTOMS, a new "one-stop shop" for CAASPP and English Language Proficiency Assessments for California (ELPAC) (September 2019)
- New statewide in-person training opportunity, the California Assessment Conference (October 2019)
- Transition from the Interim Assessment Reporting System and the Online Reporting System (ORS) to the new California Educator Reporting System (CERS), which will eventually become the "one-stop shop" for interim and summative assessment results for all CAASPP and ELPAC assessments
- Updates to Smarter Balanced Interim Assessments resources, including the 2019–2020 Interim Assessments Overview, *2019–2020 Interim Assessment User Guide*, *Interim Assessment Video Series*, and CERS Sandbox training tool
- First release of operational test questions (more than 100) from the Smarter Balanced Summative Assessments in ELA and math, as well as annotated anchor items (February 2020)
- Optional access to Smarter Balanced ELA and math Interim Assessments for use in distance learning (April 2020)
- Launch of Smarter Balanced Assessment Consortium's Tools for Teachers, and decommissioning of the Digital Library (May 2020)

HumRRO researchers engaged in a number of activities to be fully aware of these updates to the CAASPP System to understand how schools and LEAs might learn about and make use of each of the new enhancements. HumRRO researchers subscribed to the Assessment Spotlight, CDE's weekly email to educators from kindergarten to grade twelve. Launched on July 5, 2018, this publication includes information about CAASPP as well as the ELPAC. HumRRO's project management

team participated in biweekly calls with the evaluation contract monitor, Science Program staff, and DL and IA Liaison. HumRRO also attended the semiannual planning meetings conducted by the CAASPP testing contractor. Researchers reviewed new publicly available online information and attended educator training sessions supported by the CDE to understand how the updates to CAASPP components were presented to California teachers, administrators, and district staff.

For alignment of the CAST and CAA for Science, HumRRO's study designs included extensive review of each test's item and test form development documentation from CDE's testing contractor, Educational Testing Service (ETS). That work is described in chapters three and four of this report.

For year two of the Case Study, HumRRO's study design for qualitative data collection required current knowledge of the Smarter Balanced summative and interim assessments and reporting systems, as well as the Digital Library (replaced with Tools for Teachers in June 2020). HumRRO's project manager, the Case Study Director, and a Case Study researcher each observed the following CAASPP training sessions:

- *2019 Summer Institute — Analyzing Student Work and Using the Interim and Digital Library Systems to Inform Teaching and Learning.* HumRRO observed two of the eight sessions offered throughout the state, the July 8–9 session in Sacramento and the July 16–17 session in Los Angeles. Each two-day workshop was conducted by the Sacramento County Office of Education (SCOE) in partnership with WestEd and the CDE.

- *2019 California Assessment Conference* (October 16–18, Oakland). This inaugural statewide two-and-a-half-day conference was conducted by SCOE in partnership with the CDE. Educators from all roles (e.g., LEA CAASPP coordinators, professional development staff, and curriculum specialists) attended to learn how to use the state's comprehensive system of assessments (formative, interim, summative) to support teaching and learning in their schools and classrooms.

CDE's online training materials and in-person workshops for the Smarter Balanced System components emphasize the potential to impact teaching and learning when the CAASPP System tools are used in conjunction with each other. HumRRO's work with these resources is discussed in more depth in chapters two and three of this report.

Safeguarding Confidential Data

HumRRO fully understands the importance of adhering to policies that protect and monitor access to sensitive information, such as confidential test materials and data from focus groups, interviews, and online polling, while carrying out the independent evaluation activities. HumRRO researchers are cognizant of federal policies such as the Federal Educational Rights and Privacy Act (FERPA) as well as policies pertaining to governmental agencies in California and those specific to the CDE.

For the CAASPP evaluation, HumRRO staff security program training focused on three key areas: (a) proper administration of nondisclosure agreements and implementation of the "need to know" principle for all personnel working on the contract; (b) comprehensive training on specific security requirements related to HumRRO's CAASPP work, including but not limited to, specific data security and incident report procedures; and (c) clear explanation of pertinent laws and regulations governing—and the procedures related to protecting—the safeguarding of certain types of information relevant to the contract. Taken together, these areas of our security program ensure all procedures are administered in an efficient and effective manner.

Organization and Contents of the 2020 Evaluation Report

The remaining chapters of this report describe work completed during 2019–2020 for each research study listed in table 1.1.

- Chapter 2, "Instruction and Student Learning Case Study," presents HumRRO's methods and data collection activities conducted during year two of the study. The goals of the study were to learn how educators use the CAASPP Smarter Balanced System components (i.e., summative and interim assessments and the Digital Library) to inform ELA and mathematics instruction and student learning. HumRRO collected and analyzed extensive qualitative data about the use of the components in the specific context of a small number of LEAs and a small subset of each LEA's schools. HumRRO conducted in-person focus groups/interviews and monthly email polling with LEA and school educators, end-of-year web-based focus groups with LEA and school points of contact, and student questionnaires. The chapter provides, for each research question, the overarching themes and unique aspects discovered in the LEAs' use of Smarter Balanced System components during year two. The chapter concludes with best practices and recommendations for effective use of the Smarter Balanced components.

- Chapter 3, "California Science Test (CAST) Alignment Study," is an excerpt from the stand-alone technical report for this study. The chapter presents that report's Executive Summary, which includes the study's research questions to investigate the alignment of CAST to the CA NGSS, a summary of the methods and data collection activities completed, the alignment acceptability criteria HumRRO developed, and final outcomes of analysis of the alignment data and evaluation of CAST contractor documentation.

- Chapter 4, "California Alternate Assessment (CAA) for Science Alignment Study," is an excerpt from the stand-alone technical report for this study. The chapter presents that report's Executive Summary, which includes the study's research questions to investigate the alignment of CAA for Science to the Science Connectors and Focal Knowledge, Skills, and Abilities (FKSAs) derived from the CA NGSS, a summary of the methods and data collection activities completed, the alignment acceptability criteria HumRRO developed, and final outcomes of analysis of the alignment data and evaluation of CAA for Science contractor documentation.

- Chapter 5, “Conclusions and Recommendations,” provides (a) an overview of the three studies HumRRO completed during 2019–2020, (b) a summary of findings and conclusions reached for each study, (c) recommendations for improvement to the studied CAASPP components, and (d) planned updates to the CAASPP System that are anticipated to respond to several of the recommendations.

Chapter 2: Instruction and Student Learning Case Study

The two-year Impact on Instruction and Student Learning Case Study (hereafter, Case Study) uses a case study approach to deeply investigate and produce a richly detailed summary of the CAASPP System's impact in a modest number of local educational agencies (LEAs) and schools. The primary goal of the study was to elicit concrete examples of how and why specific CAASPP components (i.e., Smarter Balanced components for English language arts/literacy [ELA] and mathematics) were used and their impact on instruction and student learning, as well as the perceived benefits, strengths, and challenges of using the components. During the 2018–2019 school year, the first year of the study, HumRRO collaborated with seven LEAs, including one direct-funded charter school. The full, stand-alone report of year one of the study is available online (https://www.cde.ca.gov/Ta/Tg/ca/documents/caasppimpactcasestudy19.pdf). This chapter presents the activities and results of year two of the case study.

Creswell (1998) described a case study as an appropriate research approach when one is interested in the in-depth study of a "case" bounded in time or place. Patton (2015) noted that a "case" can be many different things, depending on the focus and field of study. Moss and Haertel (2016) use the label "Small N or Comparative Case Studies" (CCS) for studies with "more than one case, but typically fewer than fifty, purposively chosen to illuminate the question or phenomenon of interest. Typically, cases are chosen so as to contrast with respect to some set of key features. In CCS, within-case analyses are supplemented by cross-case comparisons, which help to support generalization."

For this study, a case was defined as an LEA that had fully implemented the CAASPP System in 2018–2019 and planned to continue implementation during the study year, 2019–2020 (see description in Selection of LEA Cases). To conduct a case study, one should gather a large amount of data to provide an in-depth picture of the "case" (Creswell, 1998). Like other forms of qualitative research, case studies tend to rely on use of inductive reasoning, rather than beginning with specific hypotheses (Creswell & Plano-Clark 2007). Consistent with these approaches, HumRRO's study methods relied on inductive reasoning guided by a set of research questions. HumRRO incorporated multiple types of data collection, as described further in this chapter, to provide an in-depth look at the implementation of CAASPP for a selection of LEAs and a sample of their schools.

The candor and thoughtfulness of study participants' responses to questions during all phases of data collection were the foundation of this study. Many of our LEAs and schools continued to participate in data collection activities even as they experienced COVID-19 related school closures requiring extraordinary efforts to move to virtual learning. HumRRO researchers express our deep gratitude for the time, collaboration, and contributions made by LEA and school staff to this important work.

This first section of this chapter describes the CAASPP components studied. The second section presents an abbreviated version of the study design and describes the recruitment and selection of LEAs and their associated schools. The detailed design of the Case Study is included in the 2018–2020 CAASPP Evaluation Plan, which is presented in the publicly available *2018 CAASPP Evaluation Report* (https://www.cde.ca.gov/ta/tg/ca/documents/caaspp18evalrpt.pdf).

The current report provides briefer descriptions of each aspect of the study design, including modifications made during implementation of the study, to give context for the reporting of findings.

The final section of this chapter presents general findings regarding CAASPP component use across all the LEAs studied this second year, organized by the research questions of the study. This section includes HumRRO's evaluation of contextual implications, common experiences, best practices, and challenges. The outcomes of year two of the Case Study will inform the CDE about successes as well as obstacles and suggest where potential future improvements can be made to increase the CAASPP System's intended utility to positively impact classroom instruction and student learning. The chapter concludes with a summary list of best practices and recommendations for further improvements.

The following appendices provide additional information about year two data collection for the study and more in-depth findings of CAASPP component use by each LEA and its study schools:

- Appendix A, *2019 Eligibility Survey*, presents the questionnaire HumRRO administered online to a subset of LEAs in the summer of 2019 for the purpose of identifying potential participants in year two of the study.

- Appendix B, *2019–2020 Case Study Data Collection Instruments,* presents three data collection instruments. First, the focus group protocol used to collect information from groups of teachers during site visits with participating schools. HumRRO's study included similar protocols with different question foci for school leaders and LEA staff. Second, a comprehensive list of monthly polling questions asked of LEA and school points of contacts (POCs) to learn about their ongoing use of summative assessments, Interim Assessments (IAs), and Digital Library (DL) resources. Third, an optional student survey administered by some participating schools to obtain student perspectives on the IAs.

- Appendix C, *2019–2020 Detailed LEA-Specific Findings from the Case Study*, provides an in-depth summary of the Case Study findings specific to each LEA and its study schools. This section highlights the LEA and school context and their experiences with each of the CAASPP Smarter Balanced components.

- Appendix D, *2019–2020 Summaries of LEA-Specific Findings from the Case Study*, provides summaries of the key points outlined in Appendix C for each of the LEAs.

CAASPP Smarter Balanced Components and Resources

The CAASPP System comprises multiple components intended to measure student performance and progress and serve as tools for increasing student learning in the classroom. This Case Study focused only on the CAASPP Smarter Balanced components for ELA and mathematics. This section gives an overview of the components and resources available to LEAs and schools during the 2019–2020 school year.

All the Smarter Balanced components were intentionally designed to align to the content and rigor of the Common Core State Standards (CCSS). A hierarchy of overall domain claims (most general level), sub-domain claims, assessment targets, and standards (most specific level) guide test development and contribute to analyzing and understanding the different types of Smarter Balanced scores. There are four sub-domain claims for ELA (reading, listening, writing, and research/inquiry) and four sub-domain claims for mathematics (concepts and procedures, problem solving, modeling and data analysis, and communicating and reasoning). Test results for mathematics collapse two of the mathematics claims (problem solving and modeling and data analysis) into one score reporting category. During the 2019–2020 school year, a new web tool, the Smarter Balanced Content Explorer, was launched to help educators make connections between their plans for CCSS-aligned classroom instruction and activities and the test development language of claims, targets, standards, and item specifications.

As the Smarter Balanced assessments and tools have evolved and the resources to support them expanded, finding information about a specific topic online can be challenging. The CDE maintains public web pages with information about the CAASPP System and links to documents, archived workshop presentations, webcasts, online manuals, and videos. There are also links to the CAASPP website, where online practice and training tests can be accessed.

The CAASPP website can also be accessed directly at caaspp.org (see figure 2.1). Educators use this site to access the test administration systems, training resources and materials, the latest CAASPP news, and updates regarding administering the CAASPP tests. The site has a search field and provides a wealth of information about the Smarter Balanced assessments, including updated user manuals (Resource Tab), a link to the new Smarter Balanced Content Explorer, and information about in-person or web-based training sessions (Training tab). The 2019–2020 Training Opportunities web page provided an at-a-glance view of summer and upcoming school year offerings, described the goals of in-person professional development sessions and their target audience (e.g., classroom teacher, CAASPP coordinator), and provided links to archived videos and webcasts of sessions and materials.

CDE's online resources and in-person workshops for the Smarter Balanced System components emphasize the potential to impact teaching and learning when the CAASPP System tools are used in conjunction with each other. Additionally, CDE training materials highlight the critical purpose of student assessment: to gather

evidence to make informed and appropriate instructional, policy, and programmatic decisions based on data. While encouraging educators to use all the free components, guidance in the various resources emphasizes there is no single best way to maximize the information provided by the CAASPP components. Instead, the importance of implementing CAASPP components in a manner that suits the context of a classroom, school, or district, along with other formative processes, is vital to the teaching and learning cycle.

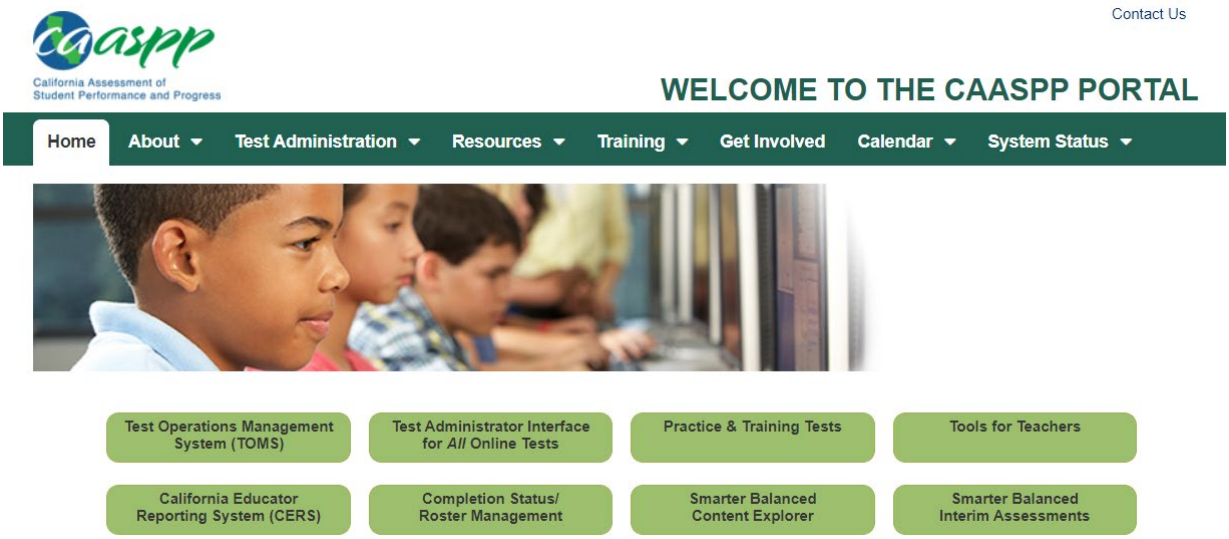

Figure 2.1 Screenshot of the home page of the CAASPP website.

Smarter Balanced Summative Assessments

The summative assessments, delivered online to students in grades three through eight and eleven, are the only Smarter Balanced component required for use in a standardized manner by all California public schools, including charter schools, in a typical academic year. [2] The summative assessments "accurately describe both student achievement (how much students know at the end of the year) and student growth (how much students have improved since the previous year) to inform program evaluation and school, district, and state accountability systems" (https://www.smarterbalanced.org/assessments/).

Each summative assessment includes a computer adaptive test (CAT) and a performance task (PT). The CAT includes a variety of item types such as selected response, constructed response, table, fill-in, and graphing. The PTs are extended activities that measure integration of knowledge and skills across multiple standards and typically require lengthier responses. The CDE provides access to aggregate results from the summative assessments on its public website (e.g., for students,

[2] 2019–2020 was not a typical year as California received a waiver for accountability testing from the Federal Government in 2019–2020 due to COVID-19 school closures.

parents, educators, researchers). Individual student reports are available only to LEA CAASPP coordinators and school test site coordinators and to parents or guardians and may be obtained only from the schools and districts where students were tested. LEAs and schools have access to a variety of score reports for their students in the Online Reporting System (ORS), and they may also download data from that system.

The CAASPP website offers educators detailed guidance and resources to support summative assessments, including:

- Online test administration manual
- Test administrator instructions (e.g., Quick Reference Guide, Checklist)
- Information about online calculator availability and sample calculators
- Information about non-embedded resources (e.g., translated test directions)

Smarter Balanced Interim Assessments

The IAs are not required but are available to California schools throughout the school year. Two main types of IAs in ELA and math were offered during the 2019–2020 school year, Interim Assessment Blocks (IABs) and Interim Comprehensive Assessments (ICAs).

- IABs are brief assessments (10 to 15 items) focused on small sets of assessment targets (up to eight); IABs provide detailed results for instructional purposes. In fall 2019, Smarter Balanced began using the name Focused Interim Assessment Blocks (FIABs) to identify IABs that measure a narrower scope of knowledge. Two main types of IAs in ELA and math were offered during the 2019–2020 school year, Interim Assessment Blocks (IABs) and Interim Comprehensive Assessments content (only one to three targets). As noted in chapter one, 42 new FIABs were made available for the 2019–2020 school year.

- ICAs cover the full range of targets and are built using the same blueprints as the summative assessments and provide results on the same scale. In 2019–2020 ICAs were released for administration to students in ninth and tenth grade to aid in early detection of college readiness. These ICAs are similar to the eleventh grade ICAs but with a grade-specific cut score for ninth and tenth grades.

All ICAs and some IABs include constructed response items; responses to these items are not machine scored and thus require hand scoring by educators, which is a local responsibility. The CDE's contractors provide hand scoring support to LEAs in the form of in-person training (e.g., at CAASPP Summer Institutes) as well as videos, online training guides, exemplars, and other training resources for use in a group setting of educators. Starting in the 2019–2020 school year, the Smarter Balanced Interim Assessment Reporting system was modified to include historical and future summative assessment results and was renamed to the California Educator Reporting System

(CERS) to reflect that change. IA results include group-level analysis (average scale score and distribution of scores across performance levels), group item-level analysis (proportion of students at each score point and item information, including item difficulty and the claim, target, and standard assessed), student-level analysis (item information, including depth of knowledge, and student responses), key and distractor analysis, and writing trait scores. Depending on how the IA was administered, results can be used by teachers "to identify students who have a strong grasp of the material and need enrichment activities to support expansion of their skills; group students by knowledge/skill level for differentiated instruction; and pinpoint areas to emphasize during classroom instruction" (Smarter Balanced Assessment Consortium, 2019).

The CAASPP website offers educators resources to support interim assessments, as illustrated in figure 2.2, a screen shot taken from the Interim Assessments link under the Resources tab. Selecting a green-shaded shape opens the link in a new browser window.

Figure 2.2 Interim Assessment Administration Resources in the CAASPP website.

In addition to the online resources, the CDE and its CAASPP partners also offer in-person training about IAs. As noted in chapter one of this report, HumRRO observed two sessions of the **2019 Summer Institute**, "Analyzing Student Work and Using the Interim Assessment and Digital Library Systems to Inform Teaching and Learning." The workshop gave researchers insight to the content and format of educator training, which was attended by some of Case Study participants, as well as an opportunity to learn about the latest system updates.

Following are a few highlights of the training noted by HumRRO researchers:

- This workshop consisted of eight modules, each including a mix of presentation of content and "table talk" among participants. Several participants told researchers they planned to share what they learned with others at their site.

- After an initial discussion of the importance of a balanced, comprehensive system of student assessments, modules covered the basics of each Smarter Balanced component, with a deeper focus on how to hand score item responses (including a performance task), how to access IA results in the online CERS, and how to find resources within the DL.

- The hands-on modules were particularly engaging and instructive for those who had not ever accessed these resources before or had not accessed them since additional enhancements were made to the features, functions, and filters of the online resources. One example was the new single sign-on system implemented to streamline access to all California student assessment systems. With one sign-on to MyTOMS, a user can access the Test Operations Management System (TOMS), CERS, and the DL. Teachers who had been frustrated with the prior need to log in separately to each system were very enthusiastic about the single sign-on.

- Participants were given a sneak preview of additional new online resources available during the 2019–2020 academic year including the Smarter Balanced Content Explorer, a searchable database to find item specifications that link claims and assessment targets to the content standards, and the Reporting System Sandbox, an open source demonstration site for learning about the functionality and reports available in CERS.

Smarter Balanced Digital Library

The DL provides instructional resources for educators to use during daily instruction in support of the formative assessment process. Individual resources can be accessed through a search by subject, grade level, specific CCSS or target, intended student population (e.g., English learners [ELs], students with disabilities [SWDs]), and other characteristics. Alternatively, educators can access playlists, which are collections of DL resources focused on similar content and organized by progressions of skills or understandings. Playlists and individual resources are also accessible through links in the IA Reporting System. This functionality allows teachers to be connected directly to

DL resources that target their students' needs. The DL also provides professional learning resources with teaching strategies. Smarter Balanced replaced the DL near the end of the 2019–2020 school year with a new online resource, "Tools for Teachers." This report refers to the DL that was functional throughout the period of data collection for the Case Study, although a preview of Tools for Teachers was made available to LEA staff in June 2020.

Study Design and Selection of LEA Cases

Research Questions

The Case Study addresses 13 key research questions pertaining to the CAASPP components of interest. Questions are organized into three general areas: (a) contextual questions and those pertaining to the full suite of Smarter Balanced components in the CAASPP System, the Summative Assessments, IAs, and DL of formative assessment tools; (b) questions related only to the Smarter Balanced Summative Assessments; and (c) questions related to the Smarter Balanced IAs and DL resources. The CAASPP Theory of Action (CDE, 2018a) was used as a guide to define these questions. Table 2.1 presents the research questions and the components they address. These questions serve as the organizing structure for presentation of the findings. HumRRO's investigation of the research questions was limited to collecting data from participating staff from the small sample of selected LEAs and their few selected schools.

Contextual conditions influence the implementation of policies and practices to a considerable degree, as noted in a recent literature review of interventions to support educators' use of data to guide decision making and practices (Marsh, 2012). Contextual conditions can be tied directly to use of data, such as the "capacity of the intervener" (e.g., guide or deliverer of training for data interpretation) and data properties (e.g., ease of interpreting outcomes of multiple measures). Broader contextual conditions include "leadership, organizational structure, time, [and] policy," as well as "interpersonal relationships and belief and knowledge."

HumRRO explored LEA and school context in terms of many factors—student demographic characteristics; academic achievement in ELA and mathematics; teacher turnover; class scheduling considerations; available curricular, technological, and other resources; professional development opportunities; and the role of professional learning communities (PLCs) of all types. For this evaluation, the acronym PLC is used as an umbrella term for organized small groups of teachers who meet regularly to collaboratively develop practice-based professional learning.

Table 2.1 CAASPP Components and Case Study Research Questions

CAASPP Components Addressed	Research Questions for Sampled LEAs and Schools
Summative, IABs and ICAs, DL	1. What are the characteristics and contexts of sampled schools/LEAs that have implemented the full suite of Smarter Balanced components?
Summative, IABs and ICAs, DL	2. How does implementation of Smarter Balanced components vary across schools/LEAs? What instructions and supports are provided to educators for implementing the components?
Summative, IABs and ICAs, DL	3. What aspects of Smarter Balanced components are perceived as most beneficial for improving classroom instruction and student learning across schools/LEAs?
Summative, IABs and ICAs, DL	4. What changes to the components and supporting resources do LEA and school staff believe would improve support for their use to improve classroom instruction and student learning?
Summative, IABs and ICAs, DL	5. How do educators/schools/LEAs use and integrate results from the summative, interim, and formative assessment resources for each content domain with each other and with other measures to enhance classroom instruction and student learning? What challenges are faced and how are they overcome?
Summative, IABs and ICAs, DL	6. How do students from schools that use the full suite of components perceive classroom opportunities to learn about summative assessment item types and topics?
Only Summative assessments	7. How do educators/schools/LEAs use summative assessment data to inform classroom instruction and make decisions?
Only IABs, ICAs, and DL	8. What interim assessments are used for ELA/literacy and mathematics for schools/LEAs that have implemented the full CAASPP System, and at what grade levels and frequency?
Only IABs, ICAs, and DL	9. What decision-making processes are used by educators/schools/LEAs to determine what interim assessments to use, who should administer them, and how frequently?
Only IABs, ICAs, and DL	10. To what extent have educators/schools/LEAs incorporated IABs into their classes? What, if any, classroom assessments have been replaced in the process? Why, and what are the implications?
Only IABs, ICAs, and DL	11. How do educators/schools/LEAs use information from ELA/literacy and mathematics interim assessments to track individual student progress and/or inform classroom instruction?
Only IABs, ICAs, and DL	12. How is information on student/school/LEA performance on interim assessments used at the school/LEA level to determine the effectiveness of practices and curricular materials for teaching the targeted standards?
DL	13. How is the DL used to improve classroom instruction?

LEA Sample

For the Case Study, HumRRO's goal was to identify and recruit six LEAs (including one charter school) that used all three CAASPP Smarter Balanced components (summative assessments, IAs, and DL) according to criteria developed jointly between HumRRO and CDE at the onset of the study (Hardoin, Thacker, Dvorak, Becker, 2018):

> *These LEAs should have demonstrated [during the prior school year] at least a "modest threshold" of use of both of the optional Smarter Balanced CAASPP components (a) IAs, with or without ICAs and hand scoring, and (b) the Instructional Resources of the Digital Library, with or without use of Professional Learning resources and Playlist resources. "Modest threshold" means a sufficient amount of use beyond simply investigating system features and will be defined based on Digital Library log-on data and interim assessment data provided to HumRRO. Eligible LEAs need not be the heaviest users in the state.*

In addition, HumRRO revised the definition for year two of the study to require some use of IAs to inform classroom instruction. In year one, we found many of our participating schools indicated they used IAs only, or primarily, to prepare students for the summative assessments. HumRRO intended to include one or two LEAs from year one to continue in year two; however, all LEAs that collaborated with HumRRO in year one who were invited to continue in year two declined.

After a review of 2018–2019 school year IA usage data and discussions regarding our desire for including schools with IA use to inform instruction, HumRRO identified the thresholds for LEA participation in the second year of the study and received CDE's approval for these eligibility criteria. HumRRO's cut point for IA usage required LEAs to include at least one school that administered at least 500 IABs in ELA and 500 IABs in mathematics during 2018–2019. No requirement was established for ICA administration, as ICA usage was much less extensive than IAB usage. Based on lessons learned during year one of our evaluation, we did not set a threshold requirement for DL logins. We learned that the login data did not capture every use of the DL when resources were accessed indirectly. In addition, many logins turned out to be teachers who accessed the system during professional development and never actually used the resources.

After using the IAB criteria to prescreen potential LEAs, HumRRO administered the 2019 Eligibility Survey to all LEAs that met the minimum requirements. The director of the CDE's Assessment Development and Administration Division emailed the county and district superintendents and charter school administrators of the prescreened LEAs to invite them to participate in the 2019 Eligibility Survey, encourage their LEA's response, and endorse the Case Study. The 2019 Eligibility Survey was similar in content to the 2018 survey, though we made slight modifications to ask if the LEA used IABs to inform classroom instruction beyond preparing for the summative assessments (see Appendix A). HumRRO administered the brief survey to further refine the set of eligible LEAs by collecting additional information about their CAASPP involvement including use of IABs to inform instruction, school characteristics, and willingness to

participate in the Case Study. HumRRO sent an invitation to complete the online survey to LEA CAASPP Coordinators. Table 2.2 summarizes survey respondents by LEA type (overall 33% response rate) and interest in participating in the study.

Table 2.2 2019 Eligibility Survey Invitees, Respondents, and Respondents' Interest in Study Participation

Respondent Type	Number of Invitees	Total Number of Respondents	Number of Respondents "Interested"	Number of Respondents "Possibly Interested"	Number of Respondents "Not Interested"
LEA (non-charter)	348	110	36	46	28
Charter	85	34	16	11	7
Total	433	144	52	57	35

Explanation of table contents: *Line 1 shows that we invited 348 non-charter LEAs to participate in our Eligibility Survey. Of these, 110 (or 32%) responded. Of the 110 respondents, 36 (33%) reported they would be potentially interested in participating in the Case Study, 46 (42%) reported they were possibly interested, and 28 (25%) were not interested.*

To choose cases from the eligible LEAs, HumRRO implemented the sampling plan outlined in the *2018 CAASPP Evaluation Report*. The goal was to identify LEAs that would very broadly represent the diversity of the state in terms of geographic region, student enrollment and demographics, and academic achievement. Based on results from the 2019 Eligibility Survey, HumRRO identified a list of the strongest candidates (15 districts and 7 charter schools) to recruit for participation. HumRRO submitted the list to the CDE for review and approval. Recruitment began with an email from HumRRO to the LEA CAASPP coordinator giving an overview of the study, followed by a teleconference call to discuss the data collection requirements of the study. For each participating LEA, HumRRO sought to include one elementary school, one middle school, and one high school. HumRRO did not seek a representative sample of schools from each LEA, but rather identified a sample of schools that were strong implementers of the Smarter Balanced components. HumRRO communicated with 19 LEAs to reach the target number of cases.

HumRRO encountered various challenges when recruiting the LEAs, which resulted in a staggered start of LEAs joining the study. First, multiple LEAs that met our criteria and indicated interest in participation were dealing with evacuations and school closures for parts of their LEA due to nearby wildfires and associated power outages. In addition, many LEAs indicated participation in other studies or initiatives that would make it difficult to participate in the Case Study. By the end of 2019, HumRRO had successfully recruited two LEAs, we gained participation of two additional LEAs in January 2020, and the final two LEAs joined the study in March 2020.

Each collaborating LEA signed a Memorandum of Understanding (MOU) with HumRRO, agreeing to participate in a specified set of data collection activities for the

duration of the 2019–2020 school year. The MOU identified a point of contact (POC) for the LEA, listed the participating schools, and identified a POC for each school. The MOU stated, in summary form, the key research questions the study sought to answer. The MOU also stated that the LEA and each school would receive a $900 honorarium for participating. Five LEAs and their schools accepted the funds, with each participating school given freedom in how these funds were used. One LEA declined its honorarium but provided an honorarium from the district to its participating schools. Due to the COVID-19 school closures, HumRRO loosened requirements for the final months of the study and provided additional honorariums for (a) administering and submitting student questionnaires and (b) participating in a web-based meeting to review preliminary findings. Despite the school closures and voluntary nature of the final months of our study, many schools continued to provide monthly polling data and participated in end-of-year virtual focus groups.

To preserve confidentiality and maintain anonymity, LEAs are identified only by number in this report (LEA-1 through LEA-6). Table 2.3 summarizes the characteristics of the six participating LEAs, which include one charter, in terms of academic achievement in ELA and mathematics and select student demographics. Data in the table are from 2018–2019. The table also indicates enrollment of students in the state or LEA who are in grades eligible for the CAASPP summative assessments.

As shown in table 2.3, statewide approximately 51 percent of students met or exceeded the grade-level standard in ELA and 40 percent did so for math. Our study LEAs spanned a range of achievement levels – LEA-4 had 81 percent of students who met or exceeded the grade-level standard in ELA and 78 percent for math, far exceeding the state average percentages. Whereas LEA-2 fell below the state average with 38 percent who met or exceeded the grade-level standard in ELA and 25 percent for math. Regarding the percentages of students meeting or exceeding the standards, readers should note that the California State Board of Education, and other states, adopted the Common Core State Standards (CCSS) in ELA and mathematics in 2010. The CCSS are generally considered more rigorous than California's previous standards and include some reorganization of content across grade levels. The Smarter Balanced Summative Assessments are aligned to the CCSS and first became operational in 2015, replacing paper and pencil assessments. Because of the substantive changes to the content standards and the time needed to implement them at the LEA and school level, the CDE anticipated the test would be very challenging to students in the initial years until adjustments to instruction caught up with the changes. In keeping with typical patterns following implementation of new standards, the statewide percentages of students meeting or exceeding the standards have been gradually increasing over time (from 2015 to 2019, an increase of 6.87% in ELA and 6.73% in mathematics) along with students' opportunity to learn the knowledge and skills measured by the assessment (Cal Matters, https://calmatters.org/education/k-12-education/2019/10/california-schools-test-scores-2019-achievement-gap-caaspp-smarter-balanced/).

Table 2.3 Characteristics of LEAs Participating in Case Study

Case Study LEA #	Total Enrollment	# CAASPP Eligible Students	% Met or Exceeded ELA State Standards	% Met or Exceeded Math State Standards	% SE Dis-advant aged	% SWD	% EL
LEA-1	103,194	48,480	55%	46%	58%	14%	21%
LEA-2	13,870	7,051	38%	25%	84%	9%	31%
LEA-3	48,936	24,745	40%	27%	90%	12%	24%
LEA-4	32,138	17,015	81%	78%	6%	9%	5%
LEA-5	9,782	4,953	43%	28%	64%	13%	24%
LEA-6	1,833	1,093	69%	49%	25%	8%	8%
All CA	6,186,278	3,189,956	51%	40%	61%	12%	19%

Explanation of table contents: *Line 1 shows that the LEA we labeled LEA-1 had a total enrollment (across all schools, including those not participating in the study) of 103,194 students in* 2018–2019*. Of these, 48,480 were eligible to participate in the CAASPP summative assessments. Of those who took the summative assessment, 55% met or exceeded the ELA state standards, and 46% met or exceeded the math state standards. In LEA-1, 58% of students were socioeconomically (SE) disadvantaged, 14% were SWDs, and 21% were ELs.*

The eligibility screening for threshold IAB usage in the sample was effective in predicting continued usage during the study year and identifying schools that used IABs to inform instructional decisions, as evidenced by information presented later in this chapter (e.g., table 2.4 below).

Data Collection

Based on the study design, HumRRO gathered data from various sources to describe the context and use of CAASPP components by each LEA and its study schools. Though HumRRO attempted to collect all information from all participants, this was challenging given the varying levels of LEA and school participation.

HumRRO collected the following data from extant sources:

- *Statewide assessment data*. Records of summative assessment administration results and counts of IAs administered in each content domain.
- *Demographic records*. Data with LEA characteristics, including student population, number of schools, student demographics, and achievement on summative assessments.

HumRRO generated data about LEA and school use of CAASPP components through the following activities:

- *Data from in-person or virtual visits to LEAs and schools*. Two HumRRO researchers prepared interview and focus group protocols (topic guides) and presented them to the CDE for review in advance of the first LEA site visit. HumRRO conducted two site visits in November 2019, two in January 2020, and two in March 2020. Because LEA-6 did not join the study until the end of February 2020 and the site visit was scheduled for mid-March, which coincided with COVID-19 school closures, HumRRO conducted interviews virtually for the POC and teachers in that LEA. See Appendix B for an example of the interview and focus group protocols.

- *Data from monthly polling of LEA and school POCs.* For five months of the study (December through April), HumRRO worked with POCs to gather LEA and school staff responses to one to three questions related to the use of Smarter Balanced components. HumRRO emailed a link to an online form that POCs could simply forward to their staff to distribute the questions, with HumRRO receiving the online responses. HumRRO informed the CDE of the question topics and provided an opportunity each month for the CDE to suggest additional questions. POC's encouraged LEA leaders, school leaders, and teachers to provide their multiple-choice and narrative responses within about one month. Due to the rolling start of cases in the study, some LEAs and schools received different questions in a particular month than did other LEAs and schools. The two late-starting cases (LEA-5 and LEA-6) did not receive the full set of monthly polling questions, and the two starting in January had "catch-up months" that incorporated multiple months of questions. See Appendix B for the full roster of school-level and LEA-level questions asked during the 2019–2020 school year.

- *Data from end of school year Web-based focus groups with LEA and school POCs*. One HumRRO researcher facilitated three online focus groups: one with LEA POCs, one with elementary school POCs, and one with middle/junior high school and high school POCs. A second researcher took detailed notes of LEA and school POCs' responses. The focus groups were audio-recorded.

- *Data from student questionnaires (optional activity)*. HumRRO asked each school POC if they were interested in collecting responses to an online questionnaire from their students to understand their experiences with IABs. HumRRO asked that each participating school provide the name of one or more math and/or ELA IABs they had recently administered and would like to survey students about. HumRRO prepared all materials for the student questionnaire, including (a) parent/guardian notification letter (English and Spanish versions) that described the goal of the questionnaire and offered parents/guardians the opportunity for their students to opt out of this activity; (b) guidelines for IAB and student selection; (c) series of step-by-step instructions for collecting student responses (excluding all personally identifiable information), including options to use an online form created by HumRRO or developing their own questionnaire using a

format of their choice. HumRRO emailed materials on April 27, 2020, to the school POCs who chose to participate, along with a request to collect information from students before the end of the school year. HumRRO received student responses from four schools representing four of the LEAs (LEA-1-HS, LEA-3-HS, LEA-5-ES1, and LEA-6).

Data Analysis Methods

The Case Study primarily involved collecting qualitative data through site visits, monthly POC polling, virtual end-of-year POC focus groups, and student questionnaire responses. HumRRO reviewed the data collected on an ongoing basis to inform questions asked during monthly polling and end-of-year focus groups. Prior to analyzing the qualitative data, HumRRO conducted several quality checks. First, immediately following each data collection activity (e.g., in-person or virtual interviews and focus groups), HumRRO researchers reviewed their notes against the audio-recording to verify accuracy of the contents and fill in any information gaps. HumRRO produced Word documents of the transcribed data. Second, HumRRO compiled monthly polling data and student response data in Excel files and conducted initial high-level coding within the file to provide indication of whether each polling question addressed summative assessments, IAs, the DL, or other topics. Monthly polling and student questionnaire data were collected using online forms and therefore did not require cleaning beyond compilation across LEAs (when separate forms were used) to prepare for analysis. After the quality assurance steps were completed, HumRRO analyzed all data sources concurrently and triangulated information to describe each LEA and its schools.

HumRRO used the text analysis features of the MAXQDA software package to analyze the qualitative data collected for the Case Study. MAXQDA is a software program designed to assist with qualitative and mixed methods data analysis. First, HumRRO created and applied a naming convention to identify the LEA and school associated with each source document. HumRRO then organized source documents by file type (e.g., LEA POC interview transcripts, teacher focus group transcripts, January monthly polling responses) and formatted them to facilitate importing. Next, HumRRO researchers imported the cleaned data files into MAXQDA. The Case Study director and researchers conducted reviews of the data in each document to (a) identify major themes and (b) revise codes identified during the year based on these data. For example, the researchers found most codes from year one regarding IAB use were still relevant in year two; however, they identified new codes related to use of FIABs. HumRRO also included codes to address the COVID-19 school closures. Though the research questions did not focus on this event, the school closures had a significant impact on the final months of our study and the topic provided important contextual information that impacted CAASPP component use. The full set of codes were reviewed and refined in an iterative fashion. The final coding system was incorporated into a single Excel document that included descriptions, and then imported into MAXQDA. HumRRO analysts used the coding system to mark text segments with similar content. Organizing and structuring the data gathered throughout the year allowed HumRRO to identify key content used to develop major themes regarding case study findings.

Four analysts were individually assigned to lead the data analysis for one or more of the six LEAs. Each analyst began with the same MAXQDA template file, preloaded with all source documents and the coding system. Using the template file, each analyst reviewed and coded data relevant only to their LEA. Analysts reviewed all text for their LEA and its schools. If text relevant to the research questions was identified but did not fit the existing codes, analysts identified new codes. The analysts communicated regularly about the coding process, especially to discuss the application of codes when the data were unclear.

For consistency in reporting the findings by LEA, the study director provided analysts a report template, along with guidance on where and how to address coded themes. Following the coding process, each analyst retrieved and reviewed coded segments to develop a draft summary of findings for their LEA(s). Two HumRRO researchers with first-hand involvement in collecting the data reviewed the LEA findings for accuracy, clarity, and consistency across sections. Analysts then reviewed, revised, and finalized their LEA sections. As a final check, HumRRO held data verification virtual meetings with LEA and school POCs who agreed to participate. To increase participation rates, POCs were offered an additional honorarium. POCs from five of the six LEAs participated in the meetings. HumRRO provided the POCs a summary of the findings relevant to their LEA or school and requested input on the accuracy. In addition to verifying that interpretations were accurate, we asked if any important information was missing regarding their use of CAASPP components. During these meetings we found HumRRO's data interpretations were highly accurate, and only a few minor clarifications were needed.

HumRRO's qualitative analysis process ensured data were systematically analyzed in a manner that captured all key information shared by LEAs and schools and treated information as similarly as possible across all LEAs. Each LEA's findings follow the major themes of the research questions (contextual factors, use of summative and interim assessments, and use of the DL). These detailed findings also include unique aspects about how each entity used the CAASPP System. The detailed LEA-specific findings are presented in Appendix D.

HumRRO's next step was to develop a summary for each LEA, consolidating the detailed LEA-specific findings and concisely reporting on the contextual factors, use of summative and interim assessments, and use of the DL. The summaries of LEA-specific findings are presented in Appendix E.

The final analysis step involved developing summaries of major themes across all schools and LEAs and relating them back to the key research questions. This was accomplished by reviewing each of the individual LEA-level summaries and noting common themes across the group of LEAs for each CAASPP component (i.e., summative assessments, IAs, and DL).

Overall Findings and Conclusions of the Case Study

This section summarizes the experiences of collaborating LEAs and schools, which we present as evidence to respond to the 13 Case Study research questions. HumRRO concludes this section with a list of best practices for using the CAASPP components and recommendations for improvements to the CAASPP System based on findings from the six LEAs studied.

School/LEA Context and Use of Full Suite of CAASPP Components

According to the theory of action for the CAASPP program, the Smarter Balanced components—working together to accurately assess student achievement relative to grade-level curriculum standards (i.e., the CCSS)—provide information to educators to help improve instruction and thus improve student achievement. The Case Study examined LEAs that are implementing the full system of components to explore how the theory of action for CAASPP components may be driving efforts for improving student achievement. The theory states that educators who use information from the system of components support high expectations, increase learning opportunities for students, and take advantage of curriculum and instructional materials and rich professional development resources to help effectively teach the content embodied by the standards.

1. **What are the characteristics and contexts of sampled schools/LEAs that have implemented the full suite of Smarter Balanced components?**

Although the plan was to identify a demographically diverse set of LEAs to participate in this study, the ultimate focus was to identify strong, collaborative CAASPP implementers who used IAs extensively, including to influence classroom instruction. Our sample met this description and included districts of various sizes, academic achievement, and demographic characteristics, as shown in table 2.4. We considered two of our five non-charter LEAs to be large, two medium, and one small. Across our six LEAs, three had a higher percentage of students who met or exceeded the ELA and mathematics grade-level standards than the state overall (51% ELA, 40% math), and three had a lower percentage of students who met or exceeded the standards. We included LEAs in southern, central, and northern California. Our LEAs included various student populations. For example, in LEA-4 only six percent of its students were classified as socioeconomically disadvantaged and only five percent as EL. In contrast, LEA-2 had 84 percent socioeconomically disadvantaged students and 31 percent classified as EL.

Though the LEAs chosen for the study were diverse in size, demographic location, and student population, we noted consistencies of learning context among them. For example, all the studied LEAs devoted time for professional learning communities (PLCs). They consistently reported schools had dedicated time in their schedules for collaboration. School leadership and teachers corroborated this information; they expressed having set out time to discuss assessment decisions, assessment data, and instructional planning. In addition, school staff across these schools were offered assistance or training regarding use of IAs and possibly other CAASPP components

(exact training differed by school). LEA and school leadership across the study were also similar in how they used their data, including CAASPP assessment data, for goal and/or decision-making purposes. Staff used summative assessment data to assist with LEA- and school-level annual planning and goal generation. The participating schools had good access to technology, with the majority having one laptop or tablet per student. Similarly, district and school leadership, and most teachers who provided data at all schools participating in the case study in year two, showed high regard for the quality of the content of the IAs and the value of IAs as measures of student progress toward grade-level standards in ELA and math. The schools selected for the study had used IABs for multiple years, and teachers were generally very familiar with how to administer them and report and use results.

Table 2.4 Characteristics of LEAs Participating in Case Study

Case Study LEA #	Location	Size	ELA Achievement	Math Achievement	% SE Dis-advan-taged	% SWD	% EL
LEA-1	Southern	Large	55%	46%	58%	14%	21%
LEA-2	Central	Medium	38%	25%	84%	9%	31%
LEA-3	Southern	Large	40%	27%	90%	12%	24%
LEA-4	Northern	Medium	81%	78%	6%	9%	5%
LEA-5	Northern	Small	43%	28%	64%	13%	24%
LEA-6 (charter)	Northern	Small	69%	49%	25%	8%	8%

2. **How does implementation of Smarter Balanced components vary across schools/LEAs? What instructions and supports are provided to educators for implementing the components?**

There were some consistencies across LEAs and their schools in use of Smarter Balanced components. For example, IABs were used to some degree by all schools in HumRRO's year two study, with some schools administering only one or two per subject area, and others electing to administer most or all IABs (see Appendix D for usage data by school). Summative assessment data were examined by LEA and school leadership and generally used as one piece of evidence to generate goals. Most schools indicated presenting data from the 2018–2019 academic year during a staff meeting early in the 2019–2020 academic year. The studied schools were mostly consistent in their use of the digital library. Though the LEAs and school administrators did not require its use, they made sure teachers were aware of its availability. The majority of the teachers across schools did not use DL resources because they did not find it easy to locate materials or they felt they already had sufficient resources through their curriculum or other sources.

LEA administrators offered various levels of support to their school sites. Across LEAS, staff were provided support to attend official CAASPP trainings. School administrators, CAASPP site leads, and often teachers were provided training by LEA staff. For example, LEA-1 provided an optional training annually, open to all school staff across the district, and a help desk available on an ongoing basis for technical issues. LEA-2 developed training for school site coordinators based on information learned from the CDE professional development trainings they attended, such as the California Assessment Conference. LEA-3-HS noted receiving professional development training from a local university with various topics including the CAASPP system. LEA-4 also provided annual CAASPP training to teachers; in this case the training focused primarily on integrating CAASPP tests and how to proctor the assessments. The LEA-5 CAASPP coordinator provided training in hand scoring for all ELA teachers and noted all teachers received one day of district-led professional development. In addition to professional development, all schools indicated collaboration time between teachers that included topics such as scheduling IAs and reviewing IA or summative assessment results. Teachers from LEA-6 had attended onsite or offsite CAASPP trainings on topics such as IA hand scoring, administering IAs, and the DL resources.

3. **What aspects of Smarter Balanced components are perceived as most beneficial for improving classroom instruction and student learning across schools/LEAs?**

School administrators and educators who participated in our study generally found IAs to be the most beneficial aspect of CAASPP for improving classroom instruction and understanding student learning. The IA benefits teachers and school leaders mentioned included exposing students to rigorous content and item-types, identifying gaps in student knowledge and determining what content needed to be retaught, and preparing students for the summative assessments. Many complaints regarding the IAs were not with the tests themselves, but the wish for more IAs, such as multiple forms of an existing IAB. Though teachers in year two of the Case Study almost always indicated the IAs were the most beneficial component, a school leader at LEA-5-ES2 noted the summative assessment results were most beneficial because the scores helped generate a five-year plan to focus instruction on student needs.

4. **What changes to the components and supporting resources do LEA and school staff believe would improve support for their use to improve classroom instruction and student learning?**

Many teachers and school administrators across LEAs in our Case Study indicated the desire for additional IABs, including traditional IABs and FIABs. Teachers wanted to see more than one IAB for targeted skills and standards in a content area and grade that would allow for (a) multiple standardized administrations to monitor progress toward achieving proficiency or (b) use in a nonstandardized manner during instruction, followed by use in a standardized manner to measure student knowledge at the end of a unit. Additionally, teachers would prefer having access to administering IAs earlier in the academic year. Some noted IABs were not available in their LEA until September or October 2019. At the time of the study, rostering was a labor-intensive process for some

LEAs and had to be completed prior to using IAs. Once rosters are in place, IAs are generally available throughout the year, with the exception of system downtimes.

Some indicated it would be useful to have summative assessment results provided earlier in the year. For example, the POC of LEA-5-ES2 indicated summative assessment results are highly important as they drive school-level goals, so having official results earlier would benefit their instructional planning. LEAs are permitted to use preliminary results as soon as they start coming in; however, not all school staff were aware of these data or had clear understanding how to use them.

Despite that all teachers in our study were familiar with IAs, some teachers were not aware of the Connections Playlist link through IA reports to DL resources. Similarly, there was little use of the DL across the study schools. School administrators and teachers indicated the DL was not user friendly, and sometimes lacked resources for grades or content areas. School leaders at LEA-1-HS suggested the DL focus resources specifically to students to allow them to independently improve – these resources could be based on IA performance and linked through IA student reports.

Study participants at two LEAs indicated it would be useful for the CDE or Smarter Balanced to maintain an updated summary of the latest resources and documentation. For example, a summary table on the CAASPP website that provides links to the most recent guidance and training, such as new videos or new versions of manuals. In a similar vein, several teachers noted dissatisfaction that updates to manuals or new assessment features were released after the school year had started or after the testing windows had opened.

LEAs and schools were generally satisfied with recent technology changes to the CAASPP system – for example, teachers at LEA-3-HS indicated they appreciated the updated single sign on for CAASPP, however, they felt that they could use additional training for CAASPP technology in general and for the DL specifically. In addition, technology improvements were recommended regarding the student rostering required before administering IAs. LEAs across our study conducted rostering at their central office for all schools – this process required many labor hours. LEA staff across the study would appreciate this process to be simplified. Other technology improvements recommended by study participants included: (a) LEA-5-ES2 requested the CDE shift system downtime to the weekends rather on school days so teachers have more options for scheduling IAs, (b) LEA-2-MS suggested the CDE make CAASPP technology more user-friendly so teachers are able to easily locate reports and different reporting features, (c) the LEA-3 POC suggested separating the test interface for the IAs and summative assessments to prevent teachers selecting the wrong link, and (d) the LEA-5 POC requested more timely assistance from the California Technical Assistance Center (CalTAC). Regarding requested improvement (c), the test administration system currently uses different colors for summative and interim assessments and includes warning notes to help prevent selection of an incorrect test.

5. **How do educators/schools/LEAs use and integrate results from the summative, interim, and formative assessment resources for each content domain with each other and with other measures to enhance classroom instruction and student learning? What challenges are faced and how are they overcome?**

LEAs, school administrators, and educators indicated making data-driven decisions based on student results on the summative and interim assessments, along with other classroom assessments. Summative assessments were often used at the school level or for initial guidance and goal setting for teachers, with IABs and other classroom assessment providing more day-to-day information.

Most teachers found the IABs, in conjunction with classroom unit assessments or other diagnostic assessments, helped them identify student strengths and weaknesses and use the data to guide future instruction. Teachers also described their exposure to IABs as motivation to increase the rigor of their day-to-day classroom instruction, such as the types of questions they build into their lessons. One school administrator from LEA-1 stated, "I would argue probably one of the most important aspects of implementing the interims is standardizing the rigor that exists in every classroom," with higher expectations for students in traditionally disadvantaged groups "who had not always been pushed in the past." A high school teacher noted that administering IAs throughout the academic year resulted in a big shift in teacher and student thinking.

Teachers expressed some challenge in using summative assessment results to inform classroom decisions because they received results for students who were no longer in their classrooms. In addition, teachers at LEA-4-HS found mandated IABs did not always align with their curriculum, and therefore were not valid measures of student learning. A consistent complaint among teachers at LEA-2 who found IAB results less useful was the lack of alignment between the instructional schedule and which IABs were scheduled and when by their LEA.

Few teachers in our study used DL resources, therefore, these rarely or never were incorporated with assessment results to enhance classroom instruction.

6. **How do students from schools that use the full suite of Smarter Balanced components perceive classroom opportunities to learn about summative assessment item types and topics for each content domain (ELA/literacy and mathematics)?**

HumRRO was unable to directly address this research question because students did not take the summative assessments in spring 2020. However, HumRRO invited school POCs to administer an online student questionnaire about IABs, which are widely used to help prepare students for the summative assessments. HumRRO collected data about student experiences with IABs during the 2019–2020 academic year from students at four schools (LEA-1 and LEA-3 high schools, an LEA-5 elementary school, and the middle and high school grades of charter LEA-6). The content of the survey is included in Appendix B.

Through these data we learned that most teachers communicated to their students that they used IABs to see how well students learned various skills. Table 2.5 provides a summary of select findings from the survey. Additional information about responses from students in participating LEAs, including student demographic information and data split by responses regarding ELA or math IABs, can be found in Appendix C. Approximately half the students recalled IAB results that led to their teacher reteaching certain skills. Some students offered information based on their IAB experiences regarding areas in which they needed to improve, including specific content areas (e.g., fractions, grammar) and test taking skills (e.g., slowing down, reading the questions more closely).

Table 2.5 Student Questionnaire Responses to Closed-Ended IAB Survey Questions

Student IAB Usage Variables	% Students LEA-1-HS (n=324)	% Students LEA-3-HS (n=114)	% Students LEA-5-ES (n=48)	% Students LEA-6 (n=7)
Standardized	80% (n=264)	93% (n = 106)	70% (*n* = 33)	71% (*n* = 5)
Standardized and Nonstandardized	8% (n=25)	1% (n=1)	9% (*n* = 4)	29% (*n* = 2)
Nonstandardized	12% (n=39)	6% (n=7)	21% (*n* = 10)	0%
Teacher's Goal: Find out what skills I have been taught/what skills I need to learn	22% (n=62)	15% (n=16)	17% (*n* = 7)	20% (*n* = 1)
Teacher's Goal: Practice certain skills	19% (n=55)	20% (n=22)	24% (*n* = 10)	20% (*n* = 1)
Teacher's Goal: Practice taking an online test	17% (n=49)	7% (n=8)	10% (*n* = 4)	20% (*n* = 1)
Teacher's Goal: See how well I learned certain skills	42% (n=122)	58% (n=64)	49% (*n* = 20)	40% (*n* = 2)
Data Used to Identify Gaps in Learning	38% (n=120)	39% (n=44)	37% (*n* = 18)	0%
Led to Teacher Reteaching Skills - Yes	46% (n=145)	44% (n=50)	54% (*n* = 26)	0%
Used special settings - Yes	23% (n=73)	36% (n=41)	11% (*n* = 5)	0%

Smarter Balanced Summative Assessment

One primary purpose of the Smarter Balanced summative assessments is to provide valid, reliable, and fair information about grades three to eight and high school students' ELA/literacy and mathematics achievement, with respect to the CCSS. The following research question explored how LEAs and schools used the data from the 2019 summative assessment during the 2019–2020 school year.

7. **How do educators/schools/LEAs use summative assessment data—including, but not limited to, information about student proficiency levels and progress towards college- and career-readiness—in ELA/literacy and mathematics to inform classroom instruction and make decisions?**

Our study LEAs and schools indicated using summative assessment results to assist with monitoring district- and school-wide performance and to generate goals. The principal of an LEA-5 elementary school noted working with an outside group to generate a five-year plan driven by summative assessment data. LEA-5 determined a districtwide need to focus on students with disabilities and English learners based on 2018–2019 summative assessment data. LEA-1-HS noted an increase in scores in 2018–2019 that followed a decrease in 2017–2018. The school considered actions taken in 2018–2019 that may have led to these increases and sought to continue them in 2019–2020.

Interim Assessments

One of the Professional Learning resources in the DL is called "Understanding the Smarter Balanced Interim Assessments." This excerpt from the resource describes research supporting the value of interim assessments: "While a rigorous summative assessment is important, it is insufficient to drive all of the change in teaching and learning. As shown by experiences in England and Hong Kong, interim and formative assessments are the other necessary assessment ingredients to drive teaching and learning (Darling-Hammond and Pechone, 2010). Grounded in cognitive development theory about how learning progresses across grades and competence develops over time (NRC, 2001; Pellegrino, 2006), Smarter Balanced interim assessments: (a) work in concert with the summative assessment; (b) allow for more innovative and fine-grained measurement of student progress toward the Common Core State Standards (Shepard, et al., 2007); and (c) provide diagnostic information that can help tailor instruction and guide students in their own learning efforts."

The following research questions explored several aspects of how LEAs and schools used the interim assessments during 2019–2020.

8. **What interim assessments are used for ELA/literacy and mathematics for schools/LEAs that have implemented the full CAASPP System, and at what grade levels and frequency?**

IABs were used by all schools included in the Case Study. Table 2.6 notes the number of schools that administered IABs in the state of California overall, and for each of our study LEAs. As shown, the average total number of IABs administered at schools that chose to use them across California was 1,095. Three of our studied LEAs administered more total IABs per school than the state average, and three administered fewer. For California overall and for five of our LEAs, more IABs were given in math on average compared to ELA. For the state and all LEAs, schools on average administered more standardized IABs than nonstandardized.

Table 2.6 Average Number of Smarter Balanced IABs Administered Per School, Statewide and by Case Study LEA, and by Subject Matter and Manner

	# Schools Giving IABs	Average # IABs Per School ELA and Math	Average # IABs Per School ELA	Average # IABs Per School Math	Average # Standardized IABs Per School (ELA and Math)	Average # Non-Standardized IABs Per School (ELA and Math)
All California	5,713	1,095	477	618	692	403
LEA-1	74	488	230	257	325	163
LEA-2	22	1,697	733	963	1,500	196
LEA-3	50	1,142	422	720	617	526
LEA-4	35	1,356	618	738	851	505
LEA-5	13	795	142	653	491	304
LEA-6	1	811	663	148	407	404

Explanation of table contents: *Row 1 shows that across all of California 5,713 schools administered IABs during the 2019–2020 school year. For these 5,713 schools, the average number of total IAB administrations was 1,095. Schools administering IABs in California on average gave 477 ELA IABs and 618 math IABs. They administered 692 IABs in a standardized manner and 403 in a nonstandardized manner (across math and ELA).*

Tables 2.7 through 2.9 summarize the total number of times ELA IABs were offered, by test name and grade, across all schools in our study. The table includes how many schools are included for each grade-level count. At the elementary school level, Read Informational Texts and Read Literary Texts were the most frequently offered ELA IABs. Most frequently offered at the middle school level was Read Informational Texts, and at the high school level, Listen/Interpret.

Table 2.7 Count of Opportunities to Take Specific IABs in English Language Arts, Across Elementary Schools in the Case Study

Test Name	Grade 3 (N Schools=7)	Grade 4 (N Schools=6)	Grade 5 (N Schools=7)	Totals
Brief Writes*	2	3	1	6
Editing**	3	3	4	10
Language and Vocabulary Use**	8	5	5	18
Listen/Interpret**	5	3	8	16
Performance Task*	3	1	1	5
Read Informational Texts*	6	6	9	21
Read Literary Texts*	7	8	6	21
Research	0	4	2	6
Research: Analyze Information**	0	1	1	2
Research: Interpret and Integrate Information**	0	3	3	6
Revision	2	2	3	7
Write and Revise Narratives**	2	1	1	4
Totals	38	40	44	122

* Indicates IAB includes some open-ended responses that require hand scoring, if the test is administered in standardized manner.
**Indicates Focused IAB.

Explanation of table contents: *These opportunities may have been a full class session or a session for a select group of students. Row 1 shows that for the schools in our study only, there were 2 opportunities (i.e., test sessions) for Brief Writes at grade 3, 3 opportunities at grade 4, and 1 opportunity at grade 5. Overall, across all our study schools, there were 6 opportunities to take Brief Writes in the elementary grades 3 through 5.*

Table 2.8 Count of Opportunities to Take Specific IABs in English Language Arts, Across Middle Schools in the Case Study

Test Name	(N Schools=6)	(N Schools=5)	(N Schools=7)	Totals
Brief Writes*	0	1	1	2
Editing**	2	1	N/A	3
Edit/Revise	N/A	N/A	2	2
Language and Vocabulary Use**	4	1	N/A	5
Listen/Interpret**	2	2	5	9
Performance Task*	0	1	0	1
Read Informational Texts*	6	8	7	21
Read Literary Texts*	3	7	7	17
Research	1	2	3	6
Research: Analyze Information**	0	1	0	1
Research: Interpret and Integrate Information**	1	2	1	4
Write and Revise Narratives**	0	1	0	1
Totals	19	27	26	52

* Indicates IAB includes some open-ended responses that require hand scoring, if the test is administered in standardized manner.

Explanation of table contents: *These opportunities may have been a full class session or a session for a select group of students. Row 1 shows that for the schools in our study only, there were 0 opportunities (i.e., test sessions) for Brief Writes at grade 6, 1 opportunity at grade 7, and 1 opportunity at grade 8. Overall, across all our study schools, there were 2 opportunities to take Brief Writes in the middle school grades 6 through 8.*

Table 2.9 Count of Opportunities to Take Specific IABs in English Language Arts, Across High Schools in the Case Study

Test Name	High School (N Schools=5)
Brief Writes*	2
Editing**	5
Edit/Revise	0
Language and Vocabulary Use**	4
Listen/Interpret**	6
Performance Task*	1
Read Informational Texts*	5
Read Literary Texts*	4
Research	3
Research: Analyze Information**	2
Research: Interpret and Integrate Information**	2
Revision	3
Write and Revise Narratives**	1
Totals	38

* Indicates IAB includes some open-ended responses that require hand scoring, if the test is administered in standardized manner.
**Indicates Focused IAB.

Explanation of table contents: *These opportunities may have been a full class session or a session for a select group of students. Row 1 shows that for the schools in our study only, there were 2 opportunities (i.e., test sessions) for Brief Writes in high school.*

Tables 2.10 through 2.12 summarize the number of times mathematics IABs were offered, by test name and grade. At the elementary school level, Number and Operations in Base Ten was the most frequently offered mathematics IAB. At the middle school level, it was Expressions and Equations, and at the high school level it was Algebra and Functions I and Algebra and Functions II.

Table 2.10 Count of Opportunities to Take Specific IABs in Mathematics, Across Elementary Schools in the Case Study

Test Name	Grade 3 (N Schools=6)	Grade 4 (N Schools=6)	Grade 5 (N Schools=8)	Totals
Add & Subtract with Equivalent Fractions**	N/A	N/A	5	5
Four Operations: Interpret, Represent, and Solve**	N/A	4	N/A	4
Fraction Equivalence and Ordering**	N/A	1	N/A	1
Geometry**	1	2	2	5
Measurement and Data	3	1	1	5
Multiply and Divide within 100**	5	N/A	N/A	5
Multiplication and Division: Interpret, Represent, and Solve**	1	N/A	N/A	1
Number and Operations - Fractions	0	2	9	11
Number and Operations – Fractions**	4	N/A	N/A	4
Number and Operations in Base Ten	0	10	12	22
Number and Operations in Base Ten**	9	N/A	N/A	9
Numerical Expressions**	N/A	N/A	2	2
Operations and Algebraic Thinking	10	4	6	20
Operations with Whole Numbers and Decimals**	N/A	N/A	6	6
Performance Task*	1	0	1	2
Properties of Multiplication & Division**	4	N/A	N/A	4
Totals	38	24	44	106

* Indicates IAB includes some open-ended responses that require hand scoring, if the test is administered in standardized manner.
**Indicates Focused IAB.

Explanation of table contents: *These opportunities may have been a full class session or a session for a select group of students. Row 1 shows that for the schools in our study only, there were 0 opportunities (i.e., test sessions) for Add & Subtract with Equivalent Fractions at grade 3 and 4, as there are no IABs of this type available for those grades. There were 5 opportunities at grade 5. Overall, across all our study schools, there were 5 opportunities to take Add & Subtract with Equivalent Fractions in the elementary grades 3 through 5.*

Table 2.11 Count of Opportunities to Take Specific IABs in Mathematics, Across Middle Schools in the Case Study

Test Name	Grade 6 (N Schools=6)	Grade 7 (N Schools=6)	Grade 8 (N Schools=6)	Totals
Algebraic Expressions & Equations**	N/A	2	N/A	2
Dependent & Independent Variables**	1	N/A	N/A	1
Divide Fractions by Fractions**	3	N/A	N/A	3
Expressions and Equations	6	5	6	17
Expressions and Equations I	N/A	N/A	10	10
Expressions and Equations II**	N/A	N/A	2	2
Functions**	N/A	N/A	6	6
Geometric Figures**	N/A	1	N/A	1
Geometry	N/A	0	2	2
One-Variable Expressions & Equations**	2	N/A	N/A	2
Ratios and Proportional Relationships**	7	3	N/A	10
The Number System	4	0	0	4
The Number System**	N/A	6	3	9
Totals	23	17	29	69

* Indicates IAB includes some open-ended responses that require hand scoring, if the test is administered in standardized manner.
**Indicates Focused IAB.

Explanation of table contents: *These opportunities may have been a full class session or a session for a select group of students. Row 1 shows that for the schools in our study only, there were 0 opportunities (i.e., test sessions) for Algebraic Expressions & Equations at grade 6 and 8, as there are no IABs of this type available for those grades. There were 2 opportunities at grade 7. Overall, across all our study schools there were 2 opportunities to take Algebraic Expressions & Equations in the middle school grades 6 through 8.*

Table 2.12 Count of Opportunities to Take Specific IABs in Mathematics, Across High Schools in the Case Study

Test Name	High School (N Schools=5)
Algebra and Functions I	5
Algebra and Functions II	5
Equations and Reasoning**	3
Geometry and Right Triangle Trigonometry**	4
Geometry Congruence	2
Geometry Measurement and Modeling	4
Interpreting Functions**	2
Number and Quantity**	2
Seeing Structure in Expressions/Polynomial Expressions**	4
Solve Equations & Inequalities: Linear and Exponential**	4
Solve Equations & Inequalities: Quadratic**	3
Statistics and Probability**	3
Total	41

* *Indicates IAB includes some open-ended responses that require hand scoring, if the test is administered in standardized manner.
**Indicates Focused IAB.

Explanation of table contents: *These opportunities may have been a full class session or a session for a select group of students. Row 1 shows that for the schools in our study only, there were 5 opportunities (i.e., test sessions) for Algebra and Functions I in high school.*

The statewide usage of ICAs (including only California schools administering at least one ICA) was far lower than that for IAB usage. Table 2.13 summarizes ICA use for all schools using ICAs across California, and for schools administering them within our studied LEAs. LEA-4 and LEA-6 did not administer ICAs, and LEA-3 included only two schools that administered, on average, 2 ICAs. Schools administering ICAs at LEA-1, LEA-2, and LEA-5 administered more than twice as many ICAs, on average, than schools that administered them across California overall.

Table 2.13 Average Number of Smarter Balanced ICAs Administered Per School, Statewide and by Case Study LEA

	# Schools Giving ICAs	Average # Total ICAs Per School ELA and Math	Average # ICAs Per School ELA	Average # ICAs Per School Math	Average # Standardized ICAs Per School (ELA and Math)	Average # Non-Standardized ICAs Per School (ELA and Math)
All California	860	55	26	29	41	14
LEA-1	5	132	60	72	127	6
LEA-2	4	201	60	141	189	13
LEA-3	2	2	0.5	1.5	1	1
LEA-4	0	N/A	N/A	N/A	N/A	N/A
LEA-5	1	162	110	52	132	30
LEA-6	0	N/A	N/A	N/A	N/A	N/A

Explanation of table contents: *Row 1 shows that across all of California 860 schools gave ICAs during the 2019–2020 school year. For these 860 schools, the average number of total ICA administrations was 55. Schools giving ICAs in California on average gave 26 ELA ICAs and 29 math ICAs. They gave 41 ICAs in a standardized manner and 14 in nonstandardized manner (across math and ELA).*

9. **What decision-making processes are used by educators/schools/LEAs to determine what ELA/literacy and mathematics interim assessments to use, who should administer them, and how frequently they should be administered?**

Case Study LEAs took different approaches in determining IAB administration. Three study LEAs mandated IA use to some degree, and three LEAs did not mandate IA use. High schools in our study generally did not assess twelfth grade students with IAs.

- LEA-2 mandated specific IAs for lower grade levels based on essential standards and summative assessment results, and mandated use at the high school level but permitted teachers to select whether they used the grade-level ICA or IABs.

- LEA-4 mandated IAB use and required two be administered for ELA and two for math, though they allowed teacher groups to decide which to administer and when.

- LEA-5 mandated selected IAs for each elementary grade, with input from teachers. A grade-level district action team looked at the standards and pacing guides and selected IAs for the district schedule. For the first time, the district

mandated three mathematics IAs at each elementary grade during the 2019–2020 academic year. The district did not require IAs in the secondary schools.

- LEA-1, LEA-3, and LEA-6 did not mandate IA use at the LEA level; however, the schools across all LEAs and at all levels selected for participation were generally strong users. LEA-1-ES made the school-level decision to administer most IABs to students in grades three through eight. Each grade level used PLC time to develop a schedule for when each IAB would be administered. LEA-3 and LEA-6 similarly did not have a mandate; however, teacher groups and school administration chose to administer IAs.

At all schools, classroom teachers administered IAs to their students. Though there were differences in frequency and schedule, all schools intended to administer IAs prior to the summative assessment administration. [3]

10. **To what extent have educators/schools/LEAs incorporated ELA/literacy and mathematics IABs into their classes? What, if any, classroom assessments have been replaced in the process? Why, and what are the implications?**

As indicated above, educators, schools, and LEAs had different levels of incorporating IABs. With the exception of a small number of teachers from LEAs where IABs were mandated, teachers in our study felt the administration of IABs were a worthwhile use of classroom time. Teachers were able to find time to administer other classroom assessments, including those from their curriculum and other sources, in addition to the IABs. Many noted that the IABs were more rigorous than what was available through their curriculum and required students to use deeper levels of thinking to respond to a question. The IABs in turn impacted classroom instruction because teachers were able to use the questions to guide the level of rigor they presented to their students.

11. **How do educators/schools/LEAs use information from ELA/literacy and mathematics interim assessments to track individual student progress and/or inform classroom instruction?**

All Case Study schools indicated using IABs to monitor student progress and/or inform classroom instruction to various degrees. Teachers at LEA-2 and LEA-4 noted using IABs at the beginning of a unit to help determine where students had prior learning and plan for how to best use their time to cover concepts. One school in LEA-5 noted teachers adjusted practices based on assessment data as part of an ongoing cycle of instruction, IAs, and adjustments to teaching practices. Educators across all study LEAs described the practice of reviewing as a class IAB questions that were problematic to many students; teachers often presented the items to the class and walked through the steps required to respond. Some teachers incorporated IAB questions into class warmup activities. A teacher at LEA-2 noted most students were not providing sufficient

[3] Though this was the plan in 2019–2020, the summative assessment administration was cancelled due to COVID-19.

textual evidence in their writing, so in response, the teachers allocated time three days a week to practice how to respond to a writing prompt. Similarly, teachers at LEA-3 noted using IAB rubrics for scoring written responses to demonstrate to students what was required for a quality writing response. Teachers across schools reported using IA results to identify areas of weakness in ELA and mathematics and adjusting instruction accordingly.

Though most schools in the study indicated IAB results were tracked at the student level primarily by teachers, LEA-1-ES tracked progress of all students at the school level. They had a goal to administer all IABs to third through fifth grade students. The principal pulled all student-level data into a shareable document that indicated how each student performed on each standard, based on the IAB results. Teachers used this information to identify student strengths and weaknesses and inform their instruction. At LEA-4-HS ELA teachers used IABs to identify incoming freshmen who needed reading remediation; the district reading specialists identified a grade eight IAB that worked well for this purpose.

12. **How is information on student/school/LEA performance on ELA/literacy and mathematics interim assessments used at the school/LEA level to determine the effectiveness of practices and curricular materials for teaching the targeted standards (i.e., CCSS)?**

Most schools in our study indicated they did not directly use IAs to determine the effectiveness of practices and curricular materials for teaching the CCSS. LEA-5 indicated reviewing data from district-mandated IABs and identifying teachers or teacher groups whose classes performed well. They sought out these teachers and identified effective teaching practices to share across the district. Though educators often indicated they did not use IAs to determine the effectiveness of their curricular material, educators across schools often described noting a difference in rigor and/or content between IAs and classroom curricula. Thus, teachers found that following their curriculum exactly was not necessarily sufficient, and they often supplemented with other resources for the skills required by the IAs or summative assessments, which reflected the CCSS. The principal of LEA-1-ES, for example, wanted additional classroom assessments similar to the rigor of the IAs and identified an online source to generate assessments that met this criterion.

Digital Library

13. **How is the Smarter Balanced Digital Library of formative tools used to improve classroom instruction (e.g., share information with students to help them monitor their own performance; better align instruction, curricula, and assessments)?**

Although most educators in our case study indicated they did not use the DL resources because they did not find it easy to identify useful resources or they felt they already had sufficient or better resources through their curriculum or otherwise, some teachers at LEA-5 accessed answer keys for writing tasks from the DL, having learned about it during

professional development. The teachers found these resources helpful to prepare students for the kind of writing expected on the summative assessments. Similarly, one teacher pulled answer keys for mathematics performance tasks from the DL.

Best Practices

Based on the full scope of second-year findings across the studied LEAs, HumRRO identified a sample of best practices supporting effective use of CAASPP components to improve teaching and learning. For this report, HumRRO defined a "best practice" as an approach used by participating LEAs, schools, or teachers that (a) aligns well with the intended purpose of and guidance for implementing components within the CAASPP System and (b) resulted in educators having a positive experience using the CAASPP System to inform their teaching. We believe these approaches may benefit other schools or LEAs that implement CAASPP.

- Use summative assessment data to monitor school-level performance and, in combination with other data, to identify school-wide goals.
- Use IAs as a teaching tool. For example, use IAs in a nonstandardized manner as a full class, small group, or partner exercise. Alternatively, review commonly missed items as a class.
- Use IA data to identify gaps in student understanding and determine content that should be retaught to the full class or select groups of students.
- Provide support and training at the school and LEA levels for using CAASPP resources. Teachers and staff who attended CAASPP professional development or reviewed resources available online increased their comfort level with the CAASPP components, including hand scoring of IABs and using and interpreting assessment results.
- Provide leadership guidance and encouragement for using CAASPP components while allowing grade-level or content-area professional learning communities (PLCs) flexibility regarding which IAs and DL resources to incorporate into their classrooms.
- Facilitate school-wide data discussions to ensure teachers know how to access and interpret summative assessment results, and how these data can inform instructional practices.
- Provide time and resources to support collaboration among grade-level and/or content-area PLCs to plan instruction and use interim and formative assessments effectively.

Recommendations

HumRRO reviewed the full scope of study findings based on the perspective of the participants—a small number of teachers within a small number of schools in a small number of LEAs—to develop suggestions for the CDE to consider as part of its continuous improvement of the CAASPP System. Some recommendations are already being addressed by planned changes to the system. Some of the planned changes include re-envisioned professional development opportunities for 2020–2021 to allow for online delivery given the COVID-19 circumstances.

Based on the second-year findings across the case study LEAs, we offer the following recommendations to the CDE to improve the CAASPP System and its implementation:

Recommendation 1: Continue providing training opportunities and updated online resources for LEA- and school-level staff. The trainings, CDE website resources, and CAASPP website resources are critical to helping educators throughout the state (a) accurately interpret Smarter Balanced summative and interim assessment results, (b) implement existing and new Smarter Balanced components, and (c) learn about enhancements to existing components.

Planned CAASPP System Changes:

- The CDE is modifying the previously held in-person Summer Institute to be a virtual Interim and Formative Assessment Training Series in October 2020. The training content will be organized into learning modules and will be structured as a "train-the-trainers" model. Local LEA staff, instructional coaches, and teachers on special assignment can in turn deliver materials to classroom teachers. Modules will include assessment literacy, interim assessment resources and systems, hand scoring practice on interim assessments, and formative assessment processes using Tools for Teachers. Three live webinars will cover these modules and provide additional guidance and support to local facilitators.

- The CDE will host a virtual statewide 2020 California Assessment Conference in October. The conference will be targeted to classroom educators with a theme of "Capitalizing on Assessment to Improve Teaching and Learning."

- The CDE will offer virtual math and ELA hand scoring workshops for teachers from December 2020 through April 2021. These workshops will be free of charge and include multiple school-day and after-school options.

Recommendation 2: Work with the Smarter Balanced Assessment Consortium to provide an expanded pool of ELA and mathematics IAs, particularly FIABs, and develop multiple versions of existing IAs. Teachers using the existing interim assessments find them of high quality and requested more options for tests for classroom use. Teachers would like new FIABs that assess additional targets. In addition, teachers commonly expressed the desire to have more than one version of each IAB/FIAB to allow use in a pre-test/post-test format or to allow use in a

nonstandardized manner as part of classroom instruction with one version, followed by standardized use of a second version for assessment.

Planned CAASPP System Changes:

- The Smarter Balanced Assessment Consortium plans to release approximately 90 more FIABs over the following two school years.

Recommendation 3: Use the CAASPP website to address the issues of version control and changing CAASPP component guidance to ensure educators are aware of new releases and use current resources. LEA and school staff indicated the CDE and Smarter Balanced provide guidance and a multitude of resources regarding CAASPP components; however, sometimes the periodic resource updates occur after the start of an academic year, making them less useful and creating some confusion about versions. Teachers would like to see CAASPP resources organized in a more structured manner with clear communication regarding how to identify and access the most current content.

Planned CAASPP System Changes:

- The CAASPP website will be housing online versions of manuals rather than static PDF versions. This will ensure that educators access the most current versions and can search for and more directly access different sections of each manual.

Recommendation 4: Consider adding reporting elements and resources directed toward students at the upper grade levels to inform their own learning. Teachers suggested high school students would benefit from targeted information regarding their strengths and weaknesses on the summative assessments and/or IAs, along with links to resources to help them improve in designated areas of weakness. Though this recommendation was provided prior to COVID-19 school closures, HumRRO believes it may be even more relevant with distance learning so prevalent.

Recommendation 5: Continue efforts to increase usability of online platforms. LEA and school staff appreciated the move to a single sign-on process in 2019–2020, though many believe there could be additional improvements to the platform. CAASPP coordinators found the process for creating groups of students (rostering) cumbersome, and schools without available LEA technical support had challenges obtaining student-level results. In addition, some teachers would like more access than they are currently provided by their school or LEA. Some teachers had difficulty remembering passwords and the reset process, while some students had issues with their login IDs. Some teachers had trouble finding IA or summative assessment score reports.

Planned CAASPP System Changes:

- The Online Reporting System (ORS) will be phased out and all CAASPP summative and interim reporting will be available through CERS.

- The CDE's planned integration of CAASPP data systems with LEA student information systems (SIS) from key vendors and several districts will provide for direct uploading of student data into CERS. This project will automate a mechanism that currently demands extensive manual effort and time to create rosters of students associated with specific teachers, and it will improve the process of obtaining score reports for a student cohort. LEAs or schools will be able to import intact groups into CERS from the LEA's SIS for rostering rather than needing to create a separate file with the groups.

- Recommendation 6: Seek ways to improve online access to high quality, free, CCSS-aligned formative assessment resources for school-level staff. The Smarter Balanced DL, which was disabled in May 2020 and replaced with the new Smarter Balanced Tools for Teachers website, was almost unused by study participants. While it was accessible during 2019–2020, the DL offered some valuable tools such as Connections Playlists, which link interim assessment results to teacher resources that help optimize student learning.

Planned CAASPP System Changes:

- Tools for Teachers was available for preview in June 2020 and had an official grand opening on September 30, 2020. The website is more user-friendly than the DL, includes high-quality materials that were reviewed by the State Network of Educators, [4] and includes the Interim Connections Playlists. The website will address many of the concerns with the DL: it is accessible (WCAG 2.1AA compliant), was purposefully developed to align with Smarter Balanced grade-level claims and targets, contains instructional resources embedded with formative assessment process strategies and accessibility strategies, and offers options and ideas for differentiation of and student access to content.

- 472BThe CDE is hosting a shared practices webinar, "**Using 'Tools for Teachers' to Support Learning,"** to orient educators to the new resource. The training webinar was conducted in September 2020 prior to the grand opening of the new website and available statewide to educators who register.

[4] The State Network of Educators is composed of educators from Smarter Balanced member states trained to contribute and review instructional and professional learning resources.

This page is intentionally blank.

Chapter 3: California Science Test Alignment Study

The California Science Test alignment study used document review and expert panel ratings to evaluate the alignment between the California Science Test (CAST) and the California Next Generation Science Standards (CA NGSS). Alignment studies are required as part of the federal assessment peer review process, provide validity evidence that the assessment is measuring the intended content, and inform future assessment item development. The CAST became operational in 2018–2019.

This chapter presents the Executive Summary from the *California Science Test (CAST) Alignment Study Report* summarizing the activities and results of this alignment study. The full, stand-alone report is available online (https://www.cde.ca.gov/ta/tg/ca/documents/castalignmentstudy0420.pdf).

Overview

The CAST is designed to measure student performance on the CA NGSS. Within the CA NGSS, performance expectations (PEs) are assessable statements of what students should know and be able to do. The following three major components, also referred to as dimensions, are combined to operationalize the PEs:

1. Disciplinary Core Ideas (DCIs) are the key ideas in science that have broad importance within or across multiple science or engineering disciplines. These core ideas build on each other as students progress through grade levels. The DCIs are grouped into the following domains: Earth and Space Sciences; Life Sciences; Physical Sciences; and Engineering, Technology, and the Application of Science (hereafter, Engineering).

2. Crosscutting Concepts (CCCs) help students explore connections across the four domains of science mentioned above in item 1. When these concepts, such as "cause and effect," are made explicit for students, they can help students develop a coherent and scientifically based view of the world around them.

3. Science and Engineering Practices (SEPs) describe what scientists do to investigate the natural world and what engineers do to design and build systems. The practices better explain and extend what is meant by "inquiry" in science and the range of cognitive, social, and physical practices that it requires. Students engage in practices to build, deepen, and apply their knowledge of core ideas and crosscutting concepts.

Evaluating alignment for the CAST represents a significant challenge because of the nature of the content, the organization of the content standards, and the test design. The three major components of the CA NGSS (DCIs, CCCs, and SEPs) are integrated into the three assessed science disciplines (Earth and Space Sciences, Life Sciences, and Physical Sciences). The test is designed such that students' knowledge is expected to be integrated and to accumulate to create a deep understanding of science content. Developing tests and test items that adequately sample such complex and integrated

content is especially challenging. When an item measures a single standard or concept, the alignment process is relatively straightforward. However, test development and alignment become more complex when standards are designed as interactions among statements about content.

The CAST is a computer-based, fixed-form (non-adaptive) assessment administered to students in grades five, eight, and once in high school (i.e., grades 10, 11, or 12). The CAST was field-tested in spring 2018 and administered operationally for the first time in January–July of 2019. The 2019 assessment included three segments, two of which contributed to an individual student's score. The third segment was used for field testing purposes only. This alignment study focused on "student-level alignment," analyzing items from the two operational segments used to compute student-level scores in order to collect evidence that individual student's scores should be sufficiently valid and reliable to support their intended interpretations. Minor changes were made to the CAST test design and blueprint in 2020 (adding one performance task and slightly reducing the number of discrete items), but those changes do not impact the conclusions drawn in this report.

The first step in evaluating for CAST alignment was to investigate the nature of the assessment itself: how the standards guided the development of the test items (and how the standards and items should therefore relate to one another) and the interpretations to be made from CAST scores. This component of the study is described in chapter 2 of the full, stand-alone report. HumRRO then modified traditional alignment methods to account for the test structure and design, a process in keeping with best practices in test validation that facilitates using alignment study results in an overall validity argument. This component of the study is described in chapter 3 of the full, stand-alone report.

Research Questions

Evidence of the alignment between assessments and standards is a requirement under the United States Department of Education's assessment peer review process. Alignment evidence supports the claim that students' test scores can be used to make valid inferences about student performance on the content being tested. The CDE identified several research questions to guide the alignment evidence collected. Activities conducted for the CAST Alignment Study were designed to provide information to answer the following research questions:

1. To what extent do the test design and test blueprint for the CAST support the claims to be made about student performance on the assessment?

2. To what extent does the test blueprint for the CAST represent an appropriate sampling of the content as set forth in the CA NGSS?

3. To what extent do the test forms and test items for the CAST reflect the test design and test blueprint?

4. To what extent do CAST tasks and items integrate DCIs, CCCs, and/or SEPs?

5. To what extent do test forms show balance across the science domains used for CAST scoring and reporting purposes (Earth and Space Sciences, Life Sciences, and Physical Sciences)?

6. Do the CAST items range from low to high cognitive complexity (i.e., depth of knowledge or DOK) and provide a sufficient number of items across the range of cognitive complexity?

7. How well does CAST fit the population being tested, in terms of the distribution of item difficulties within test forms and the distribution of student ability?

Review of CAST Documentation

HumRRO researchers collected and reviewed CAST design and test development materials provided by California Department of Education (CDE) and Educational Testing Service (ETS) staff, as well as information about the CAST shared with the public on the CDE website. HumRRO researchers evaluated the degree to which the CAST test design and development documentation met relevant standards from the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014; hereafter referred to as the *Testing Standards*).

First, HumRRO researchers identified specific standards from the *Testing Standards* that are directly relevant to how alignment is considered during test development. Next, researchers identified and collected the types of documentation needed to provide evidence that these standards were met. Finally, two HumRRO researchers independently reviewed all documentation and rated the extent to which each standard was met. These independent ratings were compared and discussed to reach a final consensus rating for each standard.

HumRRO developed and applied the following five-point rating scale to evaluate the degree to which the evidence for the assessment supports alignment to each standard:

1. No evidence of the Standard found in the materials.

2. Little evidence of the Standard found in the materials; less than half of the Standard was covered in the materials and/or evidence of key aspects of the Standard could not be found.

3. Some evidence of the Standard found in the materials; approximately half of the Standard was covered in the materials, including some key aspects of the Standard.

4. Evidence in the materials mostly covered the Standard.

5. Evidence in the materials fully covered all aspects of the Standard.

From the *Testing Standards*, the following eleven standards were identified for review:

- Standard 1.9. When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications and experience of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.

- Standard 1.11. When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.

- Standard 1.12. If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument for validity, similar information should be provided.

- Standard 2.3. For each total score, sub-score, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported.

- Standard 3.2. Test developers are responsible for developing tests that measure the intended construct and for minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other characteristics.

- Standard 3.9. Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs.

- Standard 4.0. Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.

- Standard 4.1. Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s).

- Standard 4.6. When appropriate to documenting the validity of test score interpretations for intended uses, relevant experts external to the testing program should review the test specifications to evaluate their appropriateness for intended uses of the test scores and fairness for intended test takers. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented.

- Standard 4.12. Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications.

- Standard 12.4. When a test is used as an indicator of achievement in an instructional domain or with respect to specified content standards, evidence of the extent to which the test samples the range of knowledge and elicits the processes reflected in the target domain should be provided. Both the tested and the target domains should be described in sufficient detail for their relationship to be evaluated. The analyses should make explicit those aspects of the target domain that the test represents, as well as those aspects that the test fails to represent.

All of the eleven identified standards were rated as fully met based on the available evidence. These results indicate that the CAST test design and development processes and procedures closely adhere to the testing standards related to alignment of assessment content to academic standards.

CAST Alignment Workshop and Outcomes

This CAST alignment workshop was designed to collect evidence of whether the CAST development process produces test forms that effectively measure the content and cognitive rigor reflected in the targeted content domain and the test blueprints. During the workshop, educators with content expertise evaluated how well the 2019 test items represent the associated content standards, the California Next Generation Science Standards (CA NGSS).

Alignment Criteria Evaluated

Alignment criteria were developed by HumRRO and reviewed by staff from the National Center for Improvement in Educational Assessment (Center for Assessment). These criteria were developed based on the documentation provided by CDE and ETS (the testing contractor), and they represent several aspects of the overall alignment of the

CAST to the CA NGSS. Failure to meet any single criterion does not indicate that the test is invalid or flawed in some essential way, only that that aspect of the assessment may need to be addressed through future item development or by other means.

Alignment criteria are grounded in the Webb alignment method (1997, 1999, 2002). The Webb method includes four major indicators to evaluate alignment. These indicators rely on judgments and statistical analyses to assess how well items on the assessment, regardless of item type and point value, match the state's content standards. The four alignment indicators are categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance-of-knowledge representation.

HumRRO drew from Webb's concepts (e.g., depth-of-knowledge) and the principles of Webb alignment criteria as the basis for developing alignment criteria specific to the CAST. Webb's criteria provided categories for creating alignment criteria more suited to three-dimensional assessments and content standards. For a full discussion of how and why the alignment criteria were created, see chapter 3 of the full, stand-alone report. HumRRO developed the following modified criteria for evaluating the CAST: **Link to Standards, DOK Adequacy, Range Adequacy,** and **Balance-of-Knowledge Correspondence (Revised for Science),** or simply **Balance**. To address the multidimensional nature of the CAST, we added a fifth criterion, **Multidimensional Adequacy**. Table 3.1 provides a description of each criterion.

Table 3.1 CAST-to-CA NGSS Alignment Criteria

Criterion	Description
Link to Standards	The percentage of items that panelists rate as directly and clearly matched to a PE, DCI, SEP, and/or CCC is calculated. The criterion is met if 50 percent or more of the items are matched to a specific PE and at least 90 percent of items are matched to at least one PE, DCI, SEP, or CCC.
DOK Adequacy	The percentage of items rated by panelists as reflecting each of Webb's DOK levels (Recall, Skill/Concept, Strategic Thinking) is calculated. The criterion is met if fewer than 10 percent of items are rated as DOK level 1 (Recall) and more than 10 percent of items are rated at DOK level 3 (Strategic Thinking).
Range Adequacy	The percentage of SEPs and/or CCCs that panelists rate as directly and clearly matched to one or more items is calculated. The criterion is met if at least 50 percent of CCCs and 50 percent of SEPs are aligned to test items (at least 4 CCCs and 4 SEPs).
Balance-of-Knowledge Correspondence (Revised for Science)	The number of items that panelists rate as directly and clearly matched to a content domain (e.g., Life Sciences), SEP, and/or CCC is calculated. Webb's balance-of-knowledge correspondence index is computed separately for each of these CA NGSS dimensions based on the total number of items that were matched to any content domain, SEP, and/or CCC and the proportion of those items that were matched to each specific content domain, SEP, and CCC. The criterion is met if the calculated balance index is 70 percent or higher for domains and dimensions.
Multidimensional Adequacy	The percentage of items that panelists rate as directly and clearly matched to at least one DCI, SEP, and/or CCC is calculated. The criterion is met if at least 90 percent of items are aligned to more than one dimension.

Alignment Workshop Methods

HumRRO conducted the CAST Alignment Study Workshop in the Sacramento area on February 28 and March 1, 2019. HumRRO worked collaboratively with the CDE to recruit and select a group of 18 educators to serve on one of three CAST alignment review panels (grade five, grade eight, and high school) during the two-day workshop.

Across the three panels, 14 California school districts were represented. Approximately 50 percent of panelists reported being a current teacher (including lead teacher), and the remaining 50 percent reported working in roles such as coordinator, specialist, program director, or superintendent. In addition to their current professional roles, all panelists reported having some level of experience with the CA NGSS. The types of experience reported ranged from teaching the standards to students to providing CA

NGSS-related training to other educators. Across the three panel groups, all panelists who provided responses reported experience teaching students from diverse socioeconomic and cultural backgrounds as well as experience teaching English learners.

HumRRO developed several data collection tools (see Appendix B of the full, stand-alone report) and adapted other materials to support the data collection process. Data collection tools included electronic spreadsheets for panelists and workshop facilitators to enter test item ratings. Support materials included copies of the CA NGSS and appendices (both paper and electronic), copies of the CAST item specifications, detailed workshop instructions for both panelists and facilitators, details on the cognitive complexity (DOK) rating categories and debriefing and evaluation forms. ETS created three online test forms for the alignment workshop (grade 5, 8 and high school) consisting of all the operational 2019 CAST items. ETS also created accounts for HumRRO researchers to securely access the items using the CAASPP Interim Assessment Viewing System (IAVS).

Alignment panelists received two rounds of training at the outset of the alignment workshop. First, the full group of panelists received general training that provided some background on alignment and a high-level description of the alignment process. Following the general training session, panelists moved into grade-level panel groups (grade 5, grade 8, and high school) and received more detailed training on the data collection (rating) processes and procedures.

After the panel-specific training presentation by the HumRRO facilitator, each panel engaged in a calibration activity using the first three items. Panelists accessed the items electronically and made their independent ratings. Panelists discussed their independent ratings and engaged in consensus discussion to come to agreement on the final item ratings of record. Once panelists had a clear understanding of the rating process and a common understanding of the rating categories, they moved on to rating the remaining operational items independently.

Item ratings were generated via the following steps:

1. Panelists reviewed test items independently and assigned ratings of:
    a. PE measured by item
    b. DCI measured by item (up to two DCIs, primary and secondary)
    c. CCC measured by item (up to two CCCs, primary and secondary)
    d. SEP measured by item (up to two SEPs, primary and secondary)
    e. Item Depth of Knowledge (DOK)
    f. Comments to clarify ratings or to provide feedback on quality of item or associated phenomenon
2. Panelists discussed their independent ratings.

3. HumRRO facilitator shared item metadata provided by ETS. Item metadata indicated the targeted PE, DCI, CCC, SEP, and cognitive complexity for each item.
4. Panelists came to consensus (or majority) ratings.
5. HumRRO facilitator recorded consensus/majority ratings.

The HumRRO facilitator recorded the final consensus (or majority) item ratings in a spreadsheet. Panelists then completed a debriefing form and a process evaluation survey before being released from the workshop. The debriefing form was designed to give panelists the opportunity to provide their individual, qualitative perspective on the quality of alignment. The evaluation survey elicited feedback about the quality of the workshop processes and procedures (see chapter 3 of the full, stand-alone report for more detail on workshop processes and procedures).

Alignment Workshop Results

Table 3.2 summarizes the alignment criteria results for item pools of the three summative assessment science tests. Across the three tests, panelists' ratings of the operational items provide strong support that the CAST is composed of multidimensional items that reflect a range of the CA NGSS. The ratings also support that the items generally reflect appropriate levels of cognitive complexity and a balance among the CA NGSS dimensions.

Table 3.2 Summary of Item Pool Results by Criterion and Grade Level

Criterion	Grade 5	Grade 8	High School
Links to Standards	Met	Met	Met
DOK Adequacy	Met	Partially met	Partially met
Range Adequacy	Met	Met	Met
Balance of Knowledge	Met	Partially met	Met
Multidimensional Adequacy	Met	Met	Met

Table 3.3 summarizes the test form alignment criteria results for the three summative assessment science tests. Similar to the item pool results, all test forms are composed of multidimensional items that reflect a range of the CA NGSS. Grade eight and high school test forms were evaluated as not fully reflecting an appropriate range of cognitive complexity levels, notably due to slightly more than 10 percent of items rated at DOK Level 1. Not all grade five and grade eight test forms were evaluated as fully reflecting an appropriate balance among the CA NGSS dimensions, though all calculated balance index values were within three points of the threshold value.

Table 3.3 Percentage of Grade-Level Forms Fully Meeting Each Criterion

Criterion	Grade 5	Grade 8	High School
Links to Standards	100	100	100
DOK Adequacy	100	0[a]	0[a]
Range Adequacy	100	100	100
Balance of Knowledge	60[b]	93[b]	100
Multidimensional Adequacy	100	100	100

[a] 100 percent of grade eight and high school forms at least partially met the DOK Adequacy criterion.
[b] 100 percent of grade five and eight forms at least partially met the Balance-of-Knowledge criterion.

Overall, the alignment workshop results provide strong support that the CAST is designed to produce aligned test forms. All test forms at all grade levels at least partially met all five *a priori* alignment criteria that were evaluated. Alignment criteria that were not fully met for all test forms include Depth of Knowledge Adequacy and Balance of Knowledge.

Forms that did not meet the Depth of Knowledge Adequacy criterion contained slightly more Level 1 DOK items than the 10 percent maximum outlined in the criterion (see chapter 3 of the full, stand-alone report for a detailed explanation of the alignment criteria applied). Note, also, that for each form, the number of Level 3 DOK items exceeded the ten percent minimum outlined. Failure to meet our proposed alignment criteria is often mitigated by demonstrating that test forms do meet goals outlined in test blueprints, which are reflective of the test's design and goals. At the time of this study, the CAST blueprints did not contain guidelines regarding the distribution of DOK levels. We recommend that such guidelines be added to the blueprint, along with a rationale for the range of items at each DOK level. Such a rationale might include, for example, that performance tasks are designed to lead students through simple to complex sense-making of the science phenomenon under investigation.

All forms that did not meet the Balance of Knowledge criterion were within three points of the minimum balance index threshold (adequate balance is at minimum 70 on a scale of 0 to 100). This is likely the reflection of a single or very small number of items being aligned to one dimension over another. The CA NGSS dimensions are designed to be integrated; the categories of each tend to overlap. It is not uncommon for experts to disagree with one another on the specific SEP and CCC codes that should be assigned to a test item. Although no formal confidence intervals around the minimum balance index have been established (in prior alignment research or in this study), the proximity of the calculated index values to the threshold suggest all test forms demonstrated a reasonable level of balance among the SEP and CCC categories.

Conclusions

This study combined documentation review and item ratings by content experts to evaluate the alignment between the California Science Test (CAST) and the California Next Generation Science Standards (CA NGSS). Here we present the conclusions reached for each of the seven research questions posed at the beginning of the study:

Research Question 1: To what extent do the test design and test blueprints for the CAST support the claims to be made about student performance on the assessment?

Review of available documentation found that the test design and test blueprints for the CAST support the conclusion that the testing contractor adhered to testing standards relevant to test-to-standards alignment. Review of operational test forms from the 2018–2019 administration support that the CAST design produces aligned test forms.

Research Question 2: To what extent does the test blueprint for the CAST represent an appropriate sampling of the content as set forth in the CA NGSS?

The CAST is designed such that its content at each grade level will rotate across years, each year sampling different content from the CA NGSS. The rotation is designed to allow CAST to address the full breadth of the CA NGSS over a three-year span. Table 3.4 compares the number of PEs that should be tested each year in order to meet the test blueprint with the number of PEs tested via the item pool in Year 1, based on expert panelists' ratings. The PEs assessed via the 2018–2019 item pool are sufficient to support the claim that the CAST is on track to address the full breadth of the CA NGSS after two additional operational administrations.

Table 3.4 Comparison of PE Needs per Administration and PEs Tested in Year 1

CAST Item Pool Grade Level	Physical Sciences PEs Needed Per Year	Physical Sciences PEs Tested in Year 1	Life Sciences PEs Needed Per Year	Life Sciences PEs Tested in Year 1	Earth & Space Sciences PEs Needed Per Year	Earth & Space Sciences PEs Tested in Year 1
Grade 5	5–6	11	4	10	4–5	9
Grade 8	6–7	13	7	14	5	10
High School	8	10	8	12	6–7	9

Research Question 3: To what extent do the CAST test forms and test items reflect the test design and test blueprints?

Based on expert panelists' ratings, the number of items linked to each content domain, science and engineering practice, and crosscutting concept align with the guidelines presented in the CAST blueprints. In only a small number of instances did the number of items rated as aligned to a particular dimension fall slightly outside of the ranges specified in the blueprint. Tables depicting these comparisons are presented in Appendix C of the full, stand-alone report.

Research Question 4: To what extent do CAST tasks and items integrate more than one disciplinary core idea, crosscutting concept, and/or science and engineering practice?

Expert reviewers found that most of the CAST items, across the grade levels, measure a performance expectation by integrating a disciplinary core idea, crosscutting concept, and/or science and engineering practice (and are therefore multidimensional). Table 3.5 summarizes the percentages of items across test forms that were rated as multidimensional. Across the grade levels, more than 90 percent of items were rated as multidimensional, and more than half of items on any test form were rated as integrating all three dimensions.

Table 3.5 Summary of Multidimensional Items by Grade Level

Grade Level of Test Forms	Range of Percentages of Items Aligned to Two or More Dimensions	Range of Percentages of Items Aligned to All Three Dimensions
Grade 5	91–93	64–80
Grade 8	91–98	88–95
High School	98–100	84–86

Research Question 5: To what extent do CAST test forms show balance across the disciplinary areas used for scoring and reporting purposes (Earth and Space Sciences, Life Sciences, and Physical Sciences)?

CAST forms across the grade levels reflect reasonable balance across the disciplinary areas used for scoring and reporting purposes (Earth and Space Sciences, Life Sciences, and Physical Sciences), as well as across the CA NGSS science and engineering practices and crosscutting concepts. This was determined by calculating Webb's balance index for each. This index takes into consideration (a) the number of content domains, SEPs, and CCCs measured by the items and (b) the proportion of items measuring each domain, SEP, or CCC. For most forms across the grade levels, an *a priori*-defined minimum index was met. For a smaller number of forms, this index was missed by only three points on a 100-point scale.

Research Question 6: Do the CAST items range from low to high cognitive complexity and provide a sufficient number of items across the range of cognitive complexity?

Expert reviewers indicated that CAST items vary in cognitive complexity, with slightly more than the *a priori* upper limit of 10 percent at Level 1 DOK but also more than the *a priori* minimum of 10 percent at Level 3 DOK.

Research Question 7: How well does CAST fit the population being tested, in terms of the distribution of item difficulties within test forms and the distribution of student ability?

Item-person maps, or Wright Maps, illustrate the correspondence between test takers' ability and the difficulty of the test items. Ideally, test items will be at an appropriate level of difficulty to measure the test takers' ability level, ensuring that the test provides information about test performance that is meaningful and useful. For example, test scores on a test in which most items are too difficult for most test takers would result in an inaccurate estimate of true achievement levels. Item-person maps for each grade level were produced by ETS. HumRRO conducted additional item mapping analyses, classifying items into achievement levels based on the score associated with having a 50 percent probability of responding correctly to an item (or receiving full points for a multi-point item). This classification represents the achievement level at which each item is providing the most information about student performance. Item-person maps and item-achievement level classification results are presented in Appendix D of the full, stand-alone report.

In the evaluation of the 2018–2019 operational administration, the item-person maps in Appendix D of the full, stand-alone report generally depict item difficulty being aligned with students' ability. For all three grades, the distribution of item difficulties generally lines up with the distribution of student ability levels. For high school, the item difficulty distribution relative to the student ability distribution has a slightly more upward shift compared to the other two grades. This indicates that the high school test has fewer items that are at a difficulty level that is comparable to students on the lower end of the ability distribution. Across grade levels and forms, item-achievement level classifications indicate that the largest percentage of items tended to be classified at Achievement Level 2, with some exceptions. In grade eight and high school, there were some forms in which a slightly higher percentage of items were rated at Achievement Level 4. This is in part due to multipoint items being classified based on the probability of earning full points (i.e., the ability level associated with having a 50% probability of getting the full two points on a two-point test item). Classifying items based on the probability of earning at least partial points (i.e., the ability level associated with having a 50% probability of getting at least one point on a two-point test item) would likely result in fewer items classified at Achievement Level 4.

Classifying items into achievement levels provides insight into how well a test form can differentiate among different levels of student performance. This is done by calculating the probability of answering each item correctly at each student ability level. Items are

then classified into achievement levels based on the student ability level associated with having a 50 percent probability of answering the item correctly. During standard setting, CAST achievement levels were set such that the largest percentage of students are expected to be classified at Achievement Level 2 based on the 2018–2019 spring operational test administration. Thus, it makes sense that a large proportion of items would be targeting students at this level. But test forms also contained items targeting the higher achievement levels, and, to a lesser extent, Level 1 Achievement, thus providing information about student performance at all levels. It is important to note that California educators are still developing strategies for teaching the CA NGSS in the classroom. As students have more opportunities to learn the CA NGSS, the correspondence between student ability and item difficulty is expected to shift.

Recommendations

The study results were generally very positive and do not indicate that any major changes in test development or forms construction processes and procedures are needed. We do offer one recommendation for improving the CAST blueprints:

1. **Add recommended cognitive complexity distributions to the CAST blueprints, along with a rationale for the targets set for each level.**

In lieu of adjusting the CAST blueprints themselves, establishing criteria for cognitive complexity during CAST item writing and test form construction phases will enhance alignment by clearly stating the proportions of items at each cognitive complexity level that each test form should include. This information will be helpful in ongoing evaluations of the adequacy of the item pool for building multiple test forms and for verifying that forms contain items from an appropriate range of cognitive complexity levels. These guidelines should include a rationale for each cognitive complexity level, noting why some levels are emphasized over others and how this design reflects the intent of the CA NGSS as well as the interpretation and use of CAST scores.

Chapter 4: California Alternate Assessment for Science Alignment Study

The California Alternate Assessment (CAA) for Science alignment study used document review and expert panel ratings to evaluate the alignment between the CAA for Science and the Science Core Content Connectors. The 2019–2020 CAA for Science administration was intended to be the first operational assessment. However, on March 20, 2020, all CAASPP testing was suspended due to the coronavirus disease (COVID-19) outbreak. This suspension of testing did not allow for a sufficient and representative number of students to complete the four performance tasks. Therefore, the 2020–2021 administration will be considered the first operational year, using the 2019–2020 test form.

This chapter presents the Executive Summary from the *California Alternate Assessment (CAA) for Science Alignment Study Report* summarizing the activities and results of this alignment study. The full, stand-alone report is available online (https://www.cde.ca.gov/ta/tg/ca/documents/caas19alignmentstudyrpt.pdf).

Overview

The CAA for Science is designed to measure performance on the Science Connectors. The Science Connectors are derived from the performance expectations (PEs) of the California Next Generation Science Standards (CA NGSS).

The CAA for Science is not a single end-of-year summative test but instead is designed to be administered following instruction throughout the school year. Four separate sessions, three operational and one field test, are administered each year, and each session consists of one embedded performance task (PT). Each PT addresses one science domain (i.e., Earth and Space Sciences, Life Sciences, and Physical Sciences). Administration of the CAA for Science is not tied to a typical summative assessment testing window; teachers will have discretion to administer each session when they have completed instruction on that specific domain during the school year. The students' performance on the three operational PTs will be aggregated to generate an overall science score at the conclusion of the school year. The CAA for Science is administered in grades five and eight, and once in high school. The high school assessment may be administered in grade ten, eleven, or twelve. Two Science Connectors are represented in each PT, and the five items measuring each Science Connector are expected to include two low and two medium complexity test items and one high complexity test item (numbers of score points will also vary by item). Each Science Connector has a corresponding set of five test questions prefaced by a nonscorable orienting activity designed to engage students with a science concept they were previously taught.

The first step in evaluating the alignment of the CAA for Science was to investigate the nature of the assessment itself: how the standards guided the development of the test items (and how the standards and items should therefore relate to one another) and the

interpretations to be made from CAA for Science scores. This component of the study is described in chapter 2 of the full, stand-alone report. HumRRO then modified traditional alignment methods to account for the test structure and design, a process in keeping with best practices in test validation that facilitates using alignment study results in an overall validity argument. This component of the study is described in chapter 3 of the full, stand-alone report.

Research Questions

Evidence of the alignment between assessments and standards is a requirement under the United States Department of Education's assessment peer review process. Alignment evidence supports that students' test scores can be used to make valid inferences about student performance on the content being tested. The CDE identified several research questions to guide the alignment evidence collected. Activities conducted for the CAA for Science Alignment Study were designed to provide information to answer the following research questions:

1. To what extent do the test design and test blueprint for the CAA for Science support the claims to be made about student performance on the assessment?

2. To what extent do the test forms and test items for the CAA for Science reflect the test design and test blueprint?

3. To what extent do the CAA for Science PT items link to the Science Connectors?

4. How well do the CAA for Science PT items cover the range of cognitive complexity of the Science Connectors?

Review of CAA for Science Documentation

HumRRO researchers collected and reviewed CAA for Science design and test development materials provided by California Department of Education (CDE) and Educational Testing Service (ETS) staff, as well as information about the CAA for Science shared with the public on the CDE website. HumRRO researchers evaluated the degree to which the CAA for Science test design and development documentation met relevant standards from the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014; hereafter referred to as the *Testing Standards*).

First, HumRRO researchers identified specific standards from the *Testing Standards* that are directly relevant to how alignment is considered during test development. Next, researchers identified and collected the types of documentation needed to provide evidence that these standards were met. Finally, two HumRRO researchers independently reviewed the documentation and rated the extent to which each standard was met. These independent ratings were compared and discussed to reach a final consensus rating for each standard.

HumRRO developed and applied the following five-point rating scale to evaluate the degree to which the evidence for the assessment supports alignment to each standard:

1. No evidence of the Standard found in the materials.

2. Little evidence of the Standard found in the materials; less than half of the Standard was covered in the materials and/or evidence of key aspects of the Standard could not be found.

3. Some evidence of the Standard found in the materials; approximately half of the Standard was covered in the materials, including some key aspects of the Standard.

4. Evidence in the materials mostly covered the Standard.

5. Evidence in the materials fully covered all aspects of the Standard.

From the *Testing Standards*, the following eleven standards were identified for review:

- Standard 1.9. When a validation rests in part on the opinions or decisions of expert judges, observers, or raters, procedures for selecting such experts and for eliciting judgments or ratings should be fully described. The qualifications and experience of the judges should be presented. The description of procedures should include any training and instructions provided, should indicate whether participants reached their decisions independently, and should report the level of agreement reached. If participants interacted with one another or exchanged information, the procedures through which they may have influenced one another should be set forth.

- Standard 1.11. When the rationale for test score interpretation for a given use rests in part on the appropriateness of test content, the procedures followed in specifying and generating test content should be described and justified with reference to the intended population to be tested and the construct the test is intended to measure or the domain it is intended to represent. If the definition of the content sampled incorporates criteria such as importance, frequency, or criticality, these criteria should also be clearly explained and justified.

- Standard 1.12. If the rationale for score interpretation for a given use depends on premises about the psychological processes or cognitive operations of test takers, then theoretical or empirical evidence in support of those premises should be provided. When statements about the processes employed by observers or scorers are part of the argument for validity, similar information should be provided.

- Standard 2.3. For each total score, sub-score, or combination of scores that is to be interpreted, estimates of relevant indices of reliability/precision should be reported.

- Standard 3.2. Test developers are responsible for (a) developing tests that measure the intended construct and (b) minimizing the potential for tests' being affected by construct-irrelevant characteristics, such as linguistic, communicative, cognitive, cultural, physical, or other attributes.

- Standard 3.9. Test developers and/or test users are responsible for developing and providing test accommodations, when appropriate and feasible, to remove construct-irrelevant barriers that otherwise would interfere with examinees' ability to demonstrate their standing on the target constructs.

- Standard 4.0. Tests and testing programs should be designed and developed in a way that supports the validity of interpretations of the test scores for their intended uses. Test developers and publishers should document steps taken during the design and development process to provide evidence of fairness, reliability, and validity for intended uses for individuals in the intended examinee population.

- Standard 4.1. Test specifications should describe the purpose(s) of the test, the definition of the construct or domain measured, the intended examinee population, and interpretations for intended uses. The specifications should include a rationale supporting the interpretations and uses of test results for the intended purpose(s).

- Standard 4.6. When appropriate to documenting the validity of test score interpretations for intended uses, relevant experts external to the testing program should review the test specifications to evaluate their appropriateness for intended uses of the test scores and fairness for intended test takers. The purpose of the review, the process by which the review is conducted, and the results of the review should be documented. The qualifications, relevant experiences, and demographic characteristics of expert judges should also be documented.

- Standard 4.12. Test developers should document the extent to which the content domain of a test represents the domain defined in the test specifications.

- Standard 12.4. When a test is used as an indicator of achievement in an instructional domain or with respect to specified content standards, evidence of the extent to which the test samples the range of knowledge and elicits the processes reflected in the target domain should be provided. Both the tested and the target domains should be described in sufficient detail for their relationship to be evaluated. The analyses should make explicit those aspects

of the target domain that the test represents, as well as those aspects that the test fails to represent.

All eleven standards were rated as at least partially met based on the available evidence. Most of reviewed standards (82%) were rated as mostly met (see chapter 2 of the full stand-alone report for more information). These results indicate that the test design and development processes and procedures of the CAA for Science adhere to the testing standards related to alignment of assessment content to academic standards.

CAA for Science Alignment Workshop and Outcomes

This CAA for Science alignment workshop was designed to collect evidence of whether the CAA for Science produces test forms that effectively measure the content and cognitive rigor of the targeted content domain and the test blueprint. During the workshop, educators with experience teaching students with significant cognitive disabilities and content expertise evaluated how well the 2018–2019 field test items selected for use as operational 2019–2020 items represent the associated content standards, the Science Connectors.

Alignment Criteria Evaluated

HumRRO developed alignment criteria intended to parallel those developed for the California Science Test (CAST). CAST alignment criteria were developed by HumRRO and reviewed by CDE's CAASPP Technical Advisory Group, the National Center for Improvement in Educational Assessment (Center for Assessment), and CDE staff. The CAST alignment criteria are presented in the *CAASPP CAST Alignment Study Report*.

HumRRO developed the following modified criteria for evaluating the CAA for Science: **Link to Standards, Depth of Knowledge (DOK) Adequacy, and Range Adequacy.** For a full description of the alignment criteria and discussion of how and why the alignment criteria were created, see chapter 3 of the full, stand-alone report. Failure to meet a single criterion would not indicate that the test is insufficiently aligned to generate meaningful scores, but that attention to that aspect of the test should be addressed through future item development. If several criteria were not met, we would consider this to be a signal for concern about the link between the assessment and the intended measurement construct. Table 4.1 provides a description of each criterion.

Table 4.1 CAA for Science Alignment Criteria

Criterion	Description
Link to Standards	HumRRO calculates the percentage of items panelists rate as directly and clearly matched to a Science Connector. The criterion is defined as fully met if 90% of items are matched to a Science Connector.
DOK Adequacy	HumRRO calculates the percentage of items panelists rate as reflecting each of three DOK levels (Low, Medium, and High; see Appendix B in the full stand-alone report for definitions) is calculated. The criterion is considered fully met if 25–41% of items are rated at Low Complexity, 33–50% of items are rated at Medium Complexity, and 17–33% of items are rated at High Complexity.
Range Adequacy	HumRRO calculates the percentage of items panelists rate as directly and clearly matched to one of the Focal Knowledge, Skills, and Abilities (FKSA) or to an Essential Understanding (EU). The criterion is fully met if each performance task is aligned to at least two Science Connectors and at least two EUs and one FKSA.

Alignment Workshop Methods

HumRRO conducted the CAA for Science Alignment Study Workshop in the Sacramento area on November 5 and 6, 2019. HumRRO worked collaboratively with the CDE to recruit and select a group of 18 educators to serve on one of three CAA for Science alignment review panels (grade five, grade eight, and high school) during the two-day workshop. Due to a last-minute cancellation, the high school panel included five educators rather than six.

Across the three panels, 15 California school districts were represented. Approximately 53 percent of panelists reported currently working as teachers while the remaining 47 percent reported working in roles such as inclusion specialist, instructional specialist, or program specialist. In addition to their current professional roles, 94 percent of panelists reported having some level of experience with the NGSS. The types of experience reported ranged from participating in trainings to presenting at NGSS rollouts. Across the three panels, all responding panelists reported having experience teaching students with mild to moderate and/or significant disabilities and students from diverse socioeconomic and cultural backgrounds, as well as experience teaching English learners.

HumRRO developed several data collection tools (see Appendix B of the full, stand-alone report) and adapted other materials to support the data collection process. Data collection tools included electronic spreadsheets into which panelists and workshop facilitators entered ratings for the test items that were reviewed. Support materials included copies of the (a) Connectors, (b) *Directions for Administration* (DFAs), (c) item content specifications, (d) detailed workshop instructions for both panelists and facilitators, (e) details on the cognitive complexity (DOK) rating categories, and (f) debriefing and evaluation forms.

ETS created three online test “forms” solely for use during the alignment workshop (grade five, eight and high school). These forms consisted of all the CAA for Science items that were ready for operational use in 2019−2020. ETS also created accounts for HumRRO researchers and workshop panelists to securely access the items using the CAASPP Interim Assessment Viewing System (IAVS).

Alignment panelists received two rounds of training at the outset of the alignment workshop. First, the full group of panelists received general training that provided some background on alignment and a high-level description of the alignment process. Following the general training session, panelists moved into grade-level panel groups (grade five, grade eight, and high school) and received more detailed training on the data collection (rating) processes and procedures.

After the panel-specific training presentation by the HumRRO facilitator, each panel engaged in a calibration activity using the first three items. Panelists accessed the items electronically and made their independent ratings. Panelists discussed their independent ratings and engaged in consensus discussion to come to agreement on the final item ratings of record. Once panelists had a clear understanding of the rating process and a common understanding of the rating categories, they moved on to rating the remaining operational items.

Item ratings were generated via the following steps:

1. Panelists reviewed test items independently and assigned ratings of:
    a) Connector measured by item
    b) Focal Knowledge, Skills, and Abilities (FKSAs) or Essential Understanding (EU) measured by the item
    c) Quality of the link between the item and the identified FKSA or EU
    d) Item cognitive complexity level
    e) Rating of item accessibility
    f) Comments to clarify ratings or to provide feedback on quality of item or associated phenomenon
2. Panelists discussed their independent ratings.
3. HumRRO facilitator shared item metadata. Item metadata indicated the targeted Connector, FKSA, or EU and cognitive complexity for each item.
4. Panelists came to consensus (or majority) ratings.
5. HumRRO facilitator recorded consensus/majority ratings

The HumRRO facilitator recorded the final consensus (or majority) item ratings in a spreadsheet and saved panelists’ independent ratings to a USB flash drive. Panelists then completed a debriefing form and a process evaluation survey before being released from the workshop. The debriefing form was designed to give panelists the opportunity to provide their individual, qualitative perspective on the quality of alignment.

The evaluation survey elicited feedback about the quality of the workshop processes and procedures (see chapter 3 of the full, stand-alone report for more detail on workshop processes and procedures).

Alignment Workshop Results

Table 4.2 summarizes the alignment criteria results for the three CAA for Science test item pools. Across the three tests, panelists' ratings of the operational items provide strong support that the CAA for Science consists of items that reflect the Science Connectors at a range of complexity levels.

Table 4.2 Summary of Item Pool Results by Criterion and Grade Level

Criterion	Grade Five	Grade Eight	High School
Links to Standards	Met	Met	Met
DOK Adequacy	Met	Met	Met
Range Adequacy	Met	Met	Met

Table 4.3 summarizes the by-form alignment criteria results for the three CAA for Science tests. Similar to the item pool results, all test form versions (simplified as "form" in tables) contain items that measure the Science Connectors at a range of complexity levels.

Table 4.3 Percentage of Grade-Level Forms Fully Meeting Each Criterion

Criterion	Grade Five	Grade Eight	High School
Links to Standards	100%	100%	100%
DOK Adequacy	100%	100%	50%[a]
Range Adequacy	100%	100%	100%

[a] 100% of high school form versions at least partially met the DOK Adequacy criterion.

Overall, the alignment workshop results provide strong support that the CAA for Science system produces aligned test forms. All test form versions at all grade levels at least partially met all three a priori alignment criteria. The Depth of Knowledge Adequacy criterion was not fully met for two high school test form versions; both form versions had one item more than the 41 percent acceptability threshold for Low Complexity items. Additionally, one high school form version had one item less than the 33 percent acceptability threshold for Medium Complexity items.

Conclusions

This study combined documentation review and a workshop with content experts to evaluate alignment between the California Alternate Assessment (CAA) for Science and the Science Connectors derived from the CA NGSS. Specifically, the study addressed four research questions.

Research Question 1: To what extent do the test design and test blueprint for the CAA for Science support the claims to be made about student performance on the assessment?

Review of available documentation found that the test design and test blueprint for the CAA for Science support the conclusion that the testing contractor adhered to testing standards relevant to test-to-standards alignment. Review of items that were ready for operational use in 2019–2020 supports that the CAA for Science design produces aligned test forms.

Research Question 2: To what extent do the test forms and test items for the CAA for Science reflect the test design and test blueprint?

Based on expert panelists' ratings, all performance tasks in all domains were linked to at least two Science Connectors. For two grade eight form versions, panelists identified three Science Connectors measured in the Life Sciences and Physical Sciences performance tasks. For all high school form versions, panelists identified three or more Science Connectors measured in the Life Sciences and Earth and Space Sciences performance tasks. This suggests that panelists did not find the high school performance tasks to be strongly focused on particular Science Connectors.

For nearly all grade five form versions, the number of items per task rated at each cognitive complexity level matched or was adjacent to the number outlined in the test blueprint. Similarly, for grade eight, most form versions had numbers of items rated at each level that matched or were adjacent to the blueprint guidelines. Discrepancies between panelists' ratings and blueprint guidelines were somewhat more pronounced for high school form versions, with some form versions rated as having higher numbers of low complexity Physical Sciences items and some form versions having higher numbers of medium and high complexity Life Sciences items. Tables depicting these comparisons are presented in Appendix C of the full, stand-alone report.

Research Question 3: To what extent do the CAA for Science Performance Task (PT) items link to the Science Connectors?

For all three CAA for Science tests (grade five, grade eight, and high school), all items were judged as being aligned to a Science Connector. Similarly, all performance tasks at all three grade levels measured multiple Science Connectors, Essential Understandings (EUs), and Focal Knowledge, Skills, and Abilities (FKSAs). Regardless of the version administered, every student was tested via a form that fully met the Link to Standards and Range Adequacy criteria.

Research Question 4: How well do the CAA for Science PT items cover the range of cognitive complexity of the Science Connectors?

For all three grade-level CAA for Science tests, items were rated at each of the three levels of cognitive complexity. The number of items rated at each level of cognitive complexity fell within appropriate ranges for the item pools of all three grade-level tests.

For grade five and grade eight, all test form versions included appropriate numbers of items from each cognitive complexity level. Two of the four high school test form versions had one item more than the acceptability threshold that was rated at Low Complexity. One high school test form version also had one item less than the acceptability threshold that was rated at Medium Complexity.

Chapter 5: Conclusions and Recommendations

Pursuant to California *Education Code* (*EC*) Section 60649, the Human Resources Research Organization (HumRRO) continued its independent evaluation of the California Assessment of Student Performance and Progress (CAASPP) System during the 2019–2020 academic year. This annual report covers the activities HumRRO conducted for each of the following studies:

- Instruction and Student Learning Case Study, year two
- California Science Test (CAST) Alignment Study
- California Alternate Assessment (CAA) for Science Alignment Study

This concluding chapter provides (a) an overview of the three studies HumRRO completed during 2019–2020, (b) a summary of findings and conclusions reached for each study, (c) recommendations for improvement to the studied CAASPP components, and (d) planned updates to the CAASPP System that are anticipated to respond to several of the recommendations. The final year of the Instruction and Student Learning Case Study is addressed first. The study examined use of three well-established Smarter Balanced components: Summative and Interim Assessments for ELA and mathematics, and the Digital Library (DL). Each of the two alignment studies, the CAST Alignment Study and the CAA for Science Alignment Study, are addressed next.

Instruction and Student Learning Case Study

Overview

According to the CAASPP System theory of action, the Smarter Balanced components provide information to educators to improve instruction and thus improve student achievement. The components, used in concert with each other, accurately assess student achievement relative to CDE grade-level curriculum standards. For ELA/Literacy and mathematics, the State Board of Education adopted the California Common Core State Standards (CA CCSS).

The primary goal of the two-year case study was to elicit concrete examples of how and why specific CAASPP components (i.e., Smarter Balanced components for ELA and mathematics) are used and the perceived benefits, strengths, and challenges of using the components. For each year of the case study, HumRRO collaborated with a small number of LEAs implementing Smarter Balanced Interim Assessment Blocks (IABs) for ELA or mathematics (in addition to the mandated summative assessments) to explore how the theory of action may be driving efforts to improve student achievement.

For year one of the study (2018–2019), HumRRO defined a *case* as an LEA that had at least a modest threshold of use of the IABs in 2017–2018 and planned to continue using them. HumRRO collaborated with seven LEAs during 2018–2019, including one direct-funded charter, encompassing 19 schools. The evidence collected was related

mostly to policies and practices for implementing optional CAASPP components. The small, specific group of participating LEAs, schools, and educators provided very few examples of using CAASPP components for the purpose of informing instruction or student learning. Though they cited use of IABs as important for helping students and educators prepare for the summative assessments, they relied on a mix of assessments that were locally designed, commercially purchased, or downloaded from free sources.

For year two, HumRRO defined a case as an LEA that had a robust threshold of using the IABs in 2018–2019 and whose schools included some teachers who used IABs to inform instructional decisions. We sought LEAs with at least one school that administered at least 500 ELA and 500 math IABs during the 2018–2019 school year. To relieve burden on the LEAs from the first year who were reluctant to commit to a second year and broaden coverage of districts in the state, HumRRO recruited six new LEAs for 2019–2020. The LEAs who joined the study in year two included one direct-funded charter and encompassed 15 schools. A full description of the Case Study is presented in chapter 2, including the 13 research questions; descriptions of year two LEA sample selection, data collection activities, and data analysis methods; and overall year two findings across LEAs, by research question. Appendices present in-depth and summary findings, by LEA.

Information from the second year is meaningful for the CDE and for LEAs as they consider how CAASPP components can be used in combination with other resources and what aspects might need to be improved. With the widespread school closures due to COVID-19 and the new capability of educators to administer IABs remotely as part of a distance learning approach, findings from this study may be particularly useful to support use of IABs to inform instruction during 2020–2021.

Summary of Findings

This section provides a high-level summary of the findings (across the sample of LEAs and schools in the study) associated with the use of three well-established Smarter Balanced components: summative assessments; interim assessments (IAs), which include IABs and longer Interim Comprehensive Assessments (ICAs); and the DL.

Summative Assessments

The degree to which summative assessment data were reviewed and used varied among LEAs and schools. Most school staff participating in the study reviewed data from the prior year (2018–2019) early in the first semester of the 2019–2020 school year, and many did so as a school-wide team or in professional learning communities (e.g., grade-level teams). Some delays in review of data were due to decisions made at the district level or confusion about the allowable uses of preliminary results. At a few LEAs, schools began reviewing preliminary data during spring 2019 to (a) inform site-level goals targeting improved outcomes for specified student subgroups and (b) use the data as one of several measures to help identify low-achieving students and develop intervention programs for them. This approach conforms to CDE's

encouragement of using early results to inform educational programs and support local planning around the improvement of teaching and learning.

Almost all school leaders and teachers at the elementary and middle schools reviewed grade-level results of the percentage of students who fell into each overall achievement level for ELA and mathematics and compared performance across similar districts and schools. Others also reviewed average "distance-from-three" results (i.e., the difference between the school's average scale score and the cut score for proficiency) and claim-level data broken down by achievement level. Some teachers in our study had trouble recalling anything about the prior year's summative assessment scores and thus did not describe how the results influenced instructional activities. In contrast, some schools described how summative assessment scores were a central piece of evidence for identifying annual achievement goals, and in some cases the summative assessment scores influenced instructional foci and/or the selection of IABs administered during 2019–2020.

Interim Assessments

All schools in the study used IAs in both ELA and math, except for one elementary school that did not administer any IABs in ELA. Some LEAs mandated IA use, either by indicating the minimum number of IABs and/or ICAs to be administered per subject and grade level, or by mandating the specific IABs to administer. Other LEAs allowed schools and/or individual teachers or teacher groups to make these decisions. In LEAs with mandates, teachers could administer additional IAs. Some schools or individual teachers chose to administer all or most IABs.

Many teachers cited benefits of IAs for monitoring student progress and informing instructional decisions, beyond their usefulness for preparing students for the content, rigor, item types, and technology of the summative assessments. Many teachers and school administrators across LEAs indicated the desire for additional IABs, including traditional IABs and FIABs (FIABs, which measure a narrower scope of knowledge). For standardized administrations of IABs, teachers used data from the California Educator Reporting System (CERS) to determine what specific content to reteach or review. For example, teachers identified questions with a high frequency of incorrect responses and shared results with students individually or as a class, reviewing the skills needed to solve those questions and pointing out common errors. Teachers also gave IABs in a nonstandardized manner, such as for classroom warm-ups or review activities, to practice specific areas of known student weakness. For one LEA that mandated IABs, district-wide results were used to help identify effective teaching practices to share within and across schools. At one high school, teachers used IABs to help ensure teaching is consistent with the standards. Students at that school who participated in the questionnaire reported their teachers used IABs to see how well students learned certain skills and to find out what skills they still needed to learn. The most positive perceptions about IABs were from teachers who had input into decisions about when and which IABs to give, which they found allowed for better alignment of assessments with their curriculum.

Digital Library

The study schools reported extremely limited use of the resources in the DL. Most teachers were aware of the resources and had logged directly into the DL at least once; however, at one LEA, teachers had only heard of but did not have any experience with the DL. Two LEAs provided information and training on the DL. Teachers at one school noted they accessed several resources in the DL indirectly through CERS, teachers at another school used DL resources for remediation. Many teachers noted time constraints, difficulty finding useful resources, difficulty navigating through the system, and availability of sufficient materials through their curriculum or other familiar sources as reasons for not using the DL.

Summary of Best Practices

This section provides a high-level summary of a sample of the best practices evidenced among the collaborating LEAs and schools in response to the case study research questions. The research questions addressed use of the three Smarter Balanced components studied (i.e., summative assessments, interim assessments, and the digital library). For this report, HumRRO defined a "best practice" as an approach used by participating LEAs, schools, or teachers that (a) aligned well with the intended purpose of and guidance for implementing components within the CAASPP System and (b) resulted in educators having a positive experience using the CAASPP System to inform their teaching. We believe these practices may benefit other schools or LEAs, though we acknowledge there are multiple ways to achieve the goals of the CAASPP System. Additionally, schools and LEAs need to balance approaches to meet their available resources.

Across the studied LEAs and schools, HumRRO identified the following sample of best practices used by participating LEAs for successful implementation of the Smarter Balanced components:

- Use summative assessment data to monitor school-level performance and, in combination with other data, to identify school-wide goals.
- Use IAs as a teaching tool. For example, use IAs in a nonstandardized manner as a full class, small group, or partner exercise. Alternatively, review commonly missed items as a class.
- Use IA data to identify gaps in student understanding and determine content that should be retaught to the full class or select groups of students.
- Provide support and training at the school and LEA levels for using CAASPP resources. Teachers and staff who attended CAASPP professional development or reviewed resources available online increased their comfort level with the CAASPP components, including hand scoring of IABs and using and interpreting assessment results.

- Provide leadership guidance and encouragement for using CAASPP components while allowing grade-level or content-area professional learning communities (PLCs) flexibility regarding which IAs and DL resources to incorporate into their classrooms.
- Facilitate school-wide data discussions to ensure teachers know how to access and interpret summative assessment results, and how these data can inform instructional practices.
- Provide time and resources to support collaboration among grade-level and/or content-area PLCs to plan instruction and use interim and formative assessments effectively.

Recommendations and Planned CAASPP System Changes

HumRRO reviewed the full scope of study findings based on the perspective of the participants—a small number of teachers within a small number of schools in a small number of LEAs—to develop suggestions for the CDE to consider as part of its continuous improvement of the CAASPP System.

Based on the findings across the year two case study LEAs, we offer the following recommendations. Some recommendations are already being addressed by enhancements the CDE will implement during the 2020–2021 school year. Where applicable, recommendations are followed by brief descriptions of important CAASPP System changes that will respond to areas of need described by LEA and school staff or observed by HumRRO. Some of the planned changes include re-envisioned professional development opportunities for 2020–2021 to allow for online delivery given the COVID circumstances.

Recommendation 1: Continue providing training opportunities and updated online resources for LEA- and school-level staff. The trainings, CDE website resources, and CAASPP website resources are critical to helping educators throughout the state (a) accurately interpret Smarter Balanced summative and interim assessment results, (b) implement existing and new Smarter Balanced components, and (c) learn about enhancements to existing components.

Planned CAASPP System Changes:

- The CDE is modifying the previously held in-person Summer Institute to be a virtual Interim and Formative Assessment Training Series in October 2020. The training content will be organized into learning modules and will be structured as a “train-the-trainers” model. Local LEA staff, instructional coaches, and teachers on special assignment can in turn deliver materials to classroom teachers. Modules will include assessment literacy, interim assessment resources and systems, hand scoring practice on interim assessments, and formative assessment processes using Tools for Teachers. Three live webinars will cover these modules and provide additional guidance and support to local facilitators.

- The CDE will host a virtual statewide 2020 California Assessment Conference in October. The conference will be targeted to classroom educators with a theme of "Capitalizing on Assessment to Improve Teaching and Learning."

- The CDE will offer virtual math and ELA hand scoring workshops for teachers from December 2020 through April 2021. These workshops will be free of charge and include multiple school-day and after-school options.

Recommendation 2: Work with the Smarter Balanced Assessment Consortium to provide an expanded pool of ELA and mathematics IAs, particularly FIABs, and develop multiple versions of existing IAs. Teachers using the existing interim assessments find them of high quality and requested more options for tests for classroom use. Teachers would like new FIABs that assess additional targets. In addition, teachers commonly expressed the desire to have more than one version of each IAB/FIAB to allow use in a pre-test/post-test format or to allow use in a nonstandardized manner as part of classroom instruction with one version, followed by standardized use of a second version for assessment.

Planned CAASPP System Changes:

- The Smarter Balanced Assessment Consortium plans to release approximately 90 more FIABs over the following two school years.

Recommendation 3: Use the CAASPP website to address the issues of version control and changing CAASPP component guidance to ensure educators are aware of new releases and use current resources. LEA and school staff indicated the CDE and Smarter Balanced provide guidance and a multitude of resources regarding CAASPP components; however, sometimes the periodic resource updates occur after the start of an academic year, making them less useful and creating some confusion about versions. Teachers would like to see CAASPP resources organized in a more structured manner with clear communication regarding how to identify and access the most current content.

Planned CAASPP System Changes:

- The CAASPP website will be housing online versions of manuals rather than static PDF versions. This will ensure that educators access the most current versions and can search for and more directly access different sections of each manual.

Recommendation 4: Consider adding reporting elements and resources directed toward students at the upper grade levels to inform their own learning. Teachers suggested high school students would benefit from targeted information regarding their strengths and weaknesses on the summative assessments and/or IAs (including ninth, tenth, and eleventh grade ICAs), along with links to resources to help them improve in designated areas of weakness. Though this recommendation was provided prior to COVID-19 school closures, HumRRO believes it may be even more relevant with distance learning so prevalent.

Recommendation 5: Continue efforts to increase usability of online platforms. LEA and school staff appreciated the move to a single sign-on process in 2019–2020, though many believe there could be additional improvements to the platform. CAASPP coordinators found the process for creating groups of students (rostering) cumbersome, and schools without available LEA technical support had challenges obtaining student-level results. In addition, some teachers would like more access than they are currently provided by their school or LEA. Some teachers had difficulty remembering passwords and the reset process, while some students had issues with their login IDs. Some teachers had trouble finding IA or summative assessment score reports.

Planned CAASPP System Changes:

- The Online Reporting System (ORS) will be phased out and all CAASPP summative and interim reporting will be available through CERS.

- 411BThe CDE's planned integration of CAASPP data systems with LEA student information systems (SIS) from key vendors and several districts will provide for direct uploading of student data into CERS. This project will automate a mechanism that currently demands extensive manual effort and time to create rosters of students associated with specific teachers, and it will improve the process of obtaining score reports for a student cohort. LEAs or schools will be able to import intact groups into CERS from the LEA's SIS for rostering rather than needing to create a separate file with the groups.

- Recommendation 6: Seek ways to improve online access to high quality, free, CCSS-aligned formative assessment resources for school-level staff. The Smarter Balanced DL, which was disabled in May 2020 and replaced with the new Smarter Balanced Tools for Teachers website, was almost unused by study participants. While it was accessible during 2019–2020, the DL offered some valuable tools such as Connections Playlists, which link interim assessment results to teacher resources that help optimize student learning.

Planned CAASPP System Changes:

- Tools for Teachers was available for preview in June 2020 and had an official grand opening on September 30, 2020. The website is more user-friendly than the DL, includes high-quality materials that were reviewed by the State Network of Educators, [5] and includes the Interim Connections Playlists. The website will address many of the concerns with the DL: it is accessible (WCAG 2.1AA compliant), was purposefully developed to align with Smarter Balanced grade-level claims and targets, contains instructional resources embedded with formative assessment process strategies and accessibility strategies, and offers options and ideas for differentiation of and student access to content.

[5] The State Network of Educators is composed of educators from Smarter Balanced member states trained to contribute and review instructional and professional learning resources.

- The CDE is hosting a shared practices webinar, "**Using 'Tools for Teachers' to Support Learning,"** to orient educators to the new resource. The training webinar was conducted in September 2020 prior to the grand opening of the new website and available statewide to educators who register.

California Science Test Alignment Study

Overview

The CAST is a computer-based, fixed-form (non-adaptive) assessment administered to students in grades five, eight, and once in high school (i.e., grades 10, 11, or 12). The CAST is designed to measure performance on the California Next Generation Science Standards (CA NGSS). Within the CA NGSS, performance expectations (PEs) are assessable statements of what students should know and be able to do. Three major components, also referred to as dimensions, are combined to operationalize the PEs: Disciplinary Core Ideas (DCIs), Crosscutting Concepts (CCCs), and Science and Engineering Practices (SEPs).

Evaluating alignment for the CAST represents a significant challenge because of the nature of the content, the organization of the content standards, and the test design. DCIs, CCCs, and SEPs are integrated into the three assessed science disciplines (Earth and Space Sciences, Life Sciences, and Physical Sciences). The test is designed such that students' knowledge is expected to be integrated and to accumulate to create a deep understanding of science content. Developing tests and test items that adequately sample such complex and integrated content is especially challenging.

The first step in evaluating for CAST alignment was to investigate the nature of the assessment itself: how the standards guided development of the test items (and how the standards and items should therefore relate to one another) and the interpretations to be made from CAST scores. HumRRO then modified traditional alignment methods to account for the test structure and design, a process in keeping with best practice in test validation that facilitates using alignment study results in an overall validity argument.

A summary of the activities and results of this study is presented in chapter 3, which is an excerpt from the *California Science Test (CAST) Alignment Study Report*. The full, stand-alone study report, including research questions, descriptions of data collection activities, and data analysis methods, is available online (https://www.cde.ca.gov/ta/tg/ca/documents/castalignmentstudy0420.pdf).

Summary of Findings

This section provides a high-level summary of the findings from the two major study activities. These include a review of CAST documentation and the CAST alignment workshop.

Review of CAST Documentation

HumRRO researchers collected and reviewed CAST design and test development materials provided by CDE and Educational Testing Service (ETS) staff, as well as information about the CAST shared with the public on the CDE website. HumRRO researchers evaluated the degree to which the CAST test design and development documentation met relevant standards in the *Standards for Educational and Psychological Testing* (AERA, APA & NCME, 2014; hereafter referred to as the *Testing Standards*). First, HumRRO researchers identified specific standards from the *Testing Standards* directly relevant to how alignment is considered during test development. Next, researchers identified and collected the types of documentation needed to provide evidence that these standards were met. Finally, two HumRRO researchers independently reviewed all documentation and rated the extent to which each standard was met. These independent ratings were compared and discussed to reach a final consensus rating for each standard.

All eleven identified testing standards were rated as fully met based on the available evidence. These results indicate that the CAST test design and development processes and procedures closely adhere to the testing standards related to alignment of assessment content to academic standards.

CAST Alignment Workshop

The CAST alignment workshop was designed to collect evidence of whether the CAST produces test forms that effectively measure the content and cognitive rigor reflected in the targeted content domain and the test blueprints. During the workshop, educators with content expertise evaluated how well the 2019 test items represent the associated content standards, the CA NGSS.

Alignment criteria were developed by HumRRO and reviewed by staff from the National Center for Improvement in Educational Assessment (Center for Assessment). Alignment criteria were grounded in the Webb alignment method (1997, 1999, 2002), but were modified based on the documentation provided by CDE and ETS (the testing contractor). The evaluated criteria included: **Link to Standards**, **DOK Adequacy**, **Range Adequacy**, and **Balance-of-Knowledge Correspondence (Revised for Science)**. To address the multidimensional nature of the CAST, we added a fifth criterion, **Multidimensional Adequacy**.

Across the three tests, panelists' ratings of the operational items provide strong support that the CAST is composed of multidimensional items that reflect a range of the CA NGSS. The ratings also support that the items generally reflect appropriate levels of cognitive complexity and a balance among the CA NGSS dimensions.

Similar to the item pool results, all test forms are composed of multidimensional items that reflect a range of the CA NGSS. Grade eight and high school test forms were evaluated as not fully reflecting an appropriate range of cognitive complexity levels, notably due to slightly more than 10 percent of items rated at DOK Level 1. Not all

grade five and grade eight test forms were evaluated as fully reflecting an appropriate balance among the CA NGSS dimensions, though all calculated balance index values were within three points of the threshold value.

Conclusions

This study combined documentation review and item ratings by content experts to evaluate the alignment between the CAST and the CA NGSS. Here we present the conclusions reached for each of the seven research questions posed at the beginning of the study:

Research Question 1: To what extent do the test design and test blueprints for the CAST support the claims to be made about student performance on the assessment?

Review of available documentation found that the test design and test blueprints for the CAST support the conclusion that the testing contractor adhered to testing standards relevant to test-to-standards alignment. Review of operational test forms from the 2018–2019 administration support that the CAST design produces aligned test forms.

Research Question 2: To what extent does the test blueprint for the CAST represent an appropriate sampling of the content as set forth in the CA NGSS?

The CAST is designed such that its content at each grade level will rotate across years, each year sampling different content from the CA NGSS. The rotation is designed to allow CAST to address the full breadth of the CA NGSS over a three-year span. We compared the number of PEs that should be tested each year in order to meet the test blueprint with the number of PEs tested via the item pool in year one, based on expert panelists' ratings. The PEs assessed via the 2018–2019 item pool are sufficient to support the claim that the CAST is on track to address the full breadth of the CA NGSS after two additional operational administrations.

Research Question 3: To what extent do the CAST test forms and test items reflect the test design and test blueprints?

Based on expert panelists' ratings, the number of items linked to each content domain, SEP, and CCC align with the guidelines presented in the CAST blueprints. In only a small number of instances did the number of items rated as aligned to a particular dimension fall slightly outside of the ranges specified in the blueprint.

Research Question 4: To what extent do CAST tasks and items integrate more than one disciplinary core idea, crosscutting concept, and/or science and engineering practice?

Expert reviewers found that most of the CAST items, across the grade levels, measure a PE by integrating a DCI, CCC, and/or SEP (and are therefore multidimensional). Across the grade levels, the majority of items were rated as multidimensional, and more than half of items on any test form were rated as integrating all three dimensions.

Research Question 5: To what extent do CAST test forms show balance across the disciplinary areas used for scoring and reporting purposes (Earth and Space Sciences, Life Sciences, and Physical Sciences)?

CAST forms across the grade levels reflect reasonable balance across the disciplinary areas used for scoring and reporting purposes (Earth and Space Sciences, Life Sciences, and Physical Sciences), as well as across the CA NGSS SEPs and CCCs. This was determined by calculating Webb's balance index for each disciplinary area.

Research Question 6: Do the CAST items range from low to high cognitive complexity and provide a sufficient number of items across the range of cognitive complexity?

Expert reviewers indicated that CAST items vary in cognitive complexity, with slightly more than the *a priori* limit of 10 percent at Level 1 DOK and more than the *a priori* minimum of 10 percent at Level 3 DOK.

Research Question 7: How well does CAST fit the population being tested, in terms of the distribution of item difficulties within test forms and the distribution of student ability?

Item-person maps, or Wright Maps, illustrate the correspondence between test takers' ability and the difficulty of the test items. Ideally, test items will be at an appropriate level of difficulty to measure the test takers' ability level, ensuring that the test provides information about test performance that is meaningful and useful. For all three grades, the distribution of item difficulties generally lines up with the distribution of student ability levels.

Recommendations

The study results were generally very positive and do not indicate that any major changes in test development or forms construction processes and procedures are needed. We offer one recommendation for improving the CAST blueprints.

Recommendation 1: **Add recommended cognitive complexity distributions to the CAST blueprints, along with a rationale for the targets set for each level.**

In lieu of adjusting the CAST blueprints themselves, establishing criteria for cognitive complexity during CAST item writing and test form construction phases will enhance alignment by clearly stating the proportions of items at each cognitive complexity level that each test form should include. This information will be helpful in ongoing evaluations of the adequacy of the item pool for building multiple test forms and for verifying that forms contain items from an appropriate range of cognitive complexity levels. These guidelines should include a rationale for each cognitive complexity level, noting why some levels are emphasized over others and how this design reflects the intent of the CA NGSS as well as the interpretation and use of CAST scores.

California Alternate Assessment for Science Alignment Study

Overview

The CAA for Science is designed to measure performance on the Science Connectors. The Science Connectors are derived from the performance expectations (PEs) of the CA NGSS. The CAA for Science is administered in grades five and eight, and once in high school. The high school assessment may be administered in grade ten, eleven, or twelve.

The CAA for Science is not a single end-of-year summative test but instead is designed to be administered following instruction throughout the school year. Four separate sessions, three operational and one field test, are administered each year and each session consists of one embedded performance task (PT). Each PT addresses one science domain (i.e., Earth and Space Sciences, Life Sciences, and Physical Sciences). Administration of the CAA for Science is not tied to a typical summative assessment testing window; teachers will have discretion to administer each session when they have completed instruction on that specific domain during the school year. The students' performance on the three operational PTs are aggregated to generate an overall science score at the conclusion of the school year.

The first step in evaluating the CAA for Science alignment was to investigate the nature of the assessment itself: how the content standards guided the development of the test items (and how the content standards and items should therefore relate to one another) and the interpretations to be made from CAA for Science scores. HumRRO then modified traditional alignment methods to account for the test structure and design, a process in keeping with best practice in test validation that facilitates using alignment study results in an overall validity argument.

A summary of the activities and results of this study is presented in chapter 4, which is an excerpt from the *California Alternate Assessment (CAA) for Science Alignment Study Report.* The full, stand-alone study report, including research questions, descriptions of data collection activities, and data analysis methods, is available online.

Summary of Findings

This section provides a high-level summary of the findings from the two major activities conducted for the CAA for Science Alignment Study. These included a review of CAA for Science documentation and the CAA for Science alignment workshop.

Review of CAA for Science Documentation

HumRRO researchers collected and reviewed CAA for Science design and test development materials provided by CDE and ETS staff, as well as information about the CAA for Science shared with the public on the CDE website. HumRRO researchers evaluated the degree to which the CAA for Science test design and development documentation met relevant *Testing Standards*.

First, HumRRO researchers identified specific standards from the *Testing Standards* directly relevant to how alignment is considered during test development. Next, researchers identified and collected the types of documentation needed to provide evidence that these standards were met. Finally, two HumRRO researchers independently reviewed the documentation and rated the extent to which each standard was met. These independent ratings were compared and discussed to reach a final consensus rating for each standard.

All eleven identified testing standards were rated as at least partially met based on the available evidence. Most of the reviewed standards (82%) were rated as mostly met. These results indicate that the CAA for Science test design and development processes and procedures adhere to the testing standards related to alignment of assessment content to academic standards.

CAA for Science Alignment Workshop

This CAA for Science alignment workshop was designed to collect evidence of whether the CAA for Science produces test forms that effectively measure the content and cognitive rigor reflected in the targeted content domain and the test blueprint. During the workshop, educators with experience teaching students with significant cognitive disabilities and content expertise evaluated how well the 2018–2019 field test items selected for use as operational 2019–2020 items represent the associated content standards, the Science Connectors.

HumRRO developed alignment criteria intended to parallel those developed for the CAST, but modified criteria for evaluating the CAA for Science. The evaluated criteria included: **Link to Standards**, **Depth of Knowledge (DOK) Adequacy**, and **Range Adequacy.**

Across the three grade-level tests, panelists' ratings of the operational items provide strong support that the CAA for Science consists of items that reflect the Science Connectors at a range of complexity levels. Panelists also found that all test form versions contain items that reflect the Science Connectors at a range of complexity levels. Overall, the alignment workshop results provide strong support that the CAA for Science program produces aligned test forms.

Conclusions

This study combined documentation review and a workshop with content experts to evaluate alignment between the CAA for Science and the Science Connectors derived from the CA NGSS. Specifically, the study addressed four research questions.

Research Question 1: To what extent do the test design and test blueprint for the CAA for Science support the claims to be made about student performance on the assessment?

Review of available documentation found that the test design and test blueprint for the CAA for Science support the conclusion that the testing contractor adhered to testing standards relevant to test-to-standards alignment. Review of items that were ready for operational use in 2019–2020 supports that the CAA for Science design produces aligned test forms.

Research Question 2: To what extent do the test forms and test items for the CAA for Science reflect the test design and test blueprint?

Based on expert panelists' ratings, all performance tasks in all domains were linked to at least two Science Connectors. For nearly all grade five form versions, the number of items per task rated at each cognitive complexity level matched or was adjacent to the number outlined in the test blueprint. Similarly, for grade eight, most form versions had numbers of items rated at each level that matched or were adjacent to the blueprint guidelines. Discrepancies between panelists' ratings and blueprint guidelines were somewhat more pronounced for high school form versions, with some form versions rated as having higher numbers of low complexity Physical Sciences items and some form versions having higher numbers of medium and high complexity Life Sciences items.

Research Question 3: To what extent do the CAA for Science Performance Task (PT) items link to the Science Connectors?

For all three CAA for Science tests (grade five, grade eight, and high school), all items were judged as being aligned to a Science Connector. Similarly, all PTs at all three grade levels measured multiple Science Connectors, Essential Understandings (EUs), and Focal Knowledge, Skills, and Abilities (FKSAs). Regardless of the version administered, every student was tested via a form that fully met the Link to Standards and Range Adequacy criteria.

Research Question 4: How well do the CAA for Science PT items cover the range of cognitive complexity of the Science Connectors?

For all three grade-level CAA for Science tests, items were rated at each of the three levels of cognitive complexity. The number of items rated at each level of cognitive complexity fell within appropriate ranges for the item pools of all three grade-level tests.

For grade five and grade eight, all test form versions included appropriate numbers of items from each cognitive complexity level. Two of the four high school test form versions had one item more than the acceptability threshold for Low Complexity. One high school test form version also had one item less than the acceptability threshold for Medium Complexity.

Summary and Next Steps

With the conclusion of the 2019–2020 academic year, the remaining activity for HumRRO's CAASPP independent evaluation is to prepare two additional reports. The first will be a stand-alone report on the second year of the Instruction and Student Learning Case Study. The second will be a comprehensive final report that encompasses all studies conducted during the 2018–2020 contract, including descriptions of their research designs, data collection activities, findings, and recommendations. Both reports will be delivered to the CDE by December 31, 2020. California *Education Code* (*EC*) Section 60649, which requires an independent evaluation of the CAASPP System, will become inoperative on July 1, 2021, unless an enacted statute extends this date. HumRRO has been honored to be the independent evaluator for CDE's assessment programs since 1999, contributing our objective and high-quality research efforts to support the continuous improvement of first the California High School Exit Examination and now the CAASPP System.

This page is intentionally blank.

References

Achieve. (2018). Criteria for Procuring and Evaluating High-Quality and Aligned Summative Assessments. Retrieved from https://www.achieve.org/files/Criteria03202018.pdf

California Department of Education. (2018a). Appendix A: Theory of Action for CAASPP and the Smarter Balanced Assessment System. In *California Assessment of Student Performance and Progress (CAASPP) 2018 independent evaluation report*.

California Department of Education. (2018b). *Integrating the CAASPP Tools to Create a Process of Improvement*. Presentation from 2018–19 CAASPP Institute workshops conducted across California.

Callingham, R., & Bond, T. (2006). Research in Mathematics Education and Rasch Measurement. Editorial in *Mathematics Education Research Journal*, 18, 2, 1-10.

Creswell, J.W. (1998). *Qualitative inquiry and research design: Choosing among five traditions*. Thousand Oaks, CA: Sage Publications.

Creswell, J.W., & Plano Clark, V.L. (2007). *Designing and conducting mixed methods research*. Thousand Oaks, CA: Sage Publications.

ETS. (2017). *California Science Test Blueprint*. Approved by the State Board of Education on November 8, 2017. (Revised 12/1/2017).

Darling-Hammond, L. & and Pechone, R. (2010) Developing an internationally comparable balanced assessment system that supports high-quality learning. *The National Conference on Next Generation Assessment Systems* (pp. 3–63). The Center for K–12 Assessment & Performance Management.

Dickinson, E., Thacker, A., Hardoin, M., (2020). *California Assessment of Student Performance and Progress (CAASPP) California Science Test alignment study report* (2020 No. 040). Alexandria, VA: Human Resources Research Organization. https://www.cde.ca.gov/ta/tg/ca/documents/castalignmentstudy0420.pdf

Dickinson, E., Thacker, A., (2020). *California Assessment of Student Performance and Progress (CAASPP) California Alternate Assessment for Science alignment study report* (2020 No. 046). Alexandria, VA: Human Resources Research Organization. https://www.cde.ca.gov/ta/tg/ca/documents/caas19alignmentstudyrpt.pdf

Flowers, C., Wakeman, S., Browder, D., & Karvonen, M. (2007). *Links for academic learning: An alignment protocol for alternate assessments based on alternate achievement standards.* Charlotte, NC: University of North Carolina at Charlotte. Retrieved from: http://www.naacpartners.org/LAL/documents/NAAC_AlignmentManualVer8_3.pdf

Hardoin, M. M., Thacker, A., Norman Dvorak, R., Becker, D. E. (2018). *California Assessment of Student Performance and Progress (CAASPP) independent evaluation report* (2018 No. 087). Alexandria, VA: Human Resources Research Organization. https://www.cde.ca.gov/ta/tg/ca/documents/caaspp18evalrpt.pdf

Hardoin, M. M., Dvorak, R. L., Thacker, A. A., Paulsen, J., Gribben, M. A., & Handy, K. (2019a). *California Assessment of Student Performance and Progress (CAASPP): 2019 independent evaluation report* (2019 No. 102). Alexandria, VA: Human Resources Research Organization. https://www.cde.ca.gov/ta/tg/ca/documents/caaspp19evalrpt.pdf

Hardoin, M. M., Dvorak, R. L., Paulsen, J., Gribben, M. A., & Handy, K. (2019b). *California Assessment of Student Performance and Progress (CAASPP) 2019 impact case study report* (2019 No. 103). Alexandria, VA: Human Resources Research Organization. https://www.cde.ca.gov/ta/tg/ca/documents/caasppimpactcasestudy19.pdf

Marsh, J.A. (2012). Interventions promoting educators' use of data: Research insights and gaps. — *College Record, 11*(2). 1–48.

Moss, P. A., & Haertel, E. H. (2016). Engaging methodological pluralism. In D. Gitomer and C. Bell (Eds.), Handbook of research on teaching (5th ed.), (pp. 127–247). Washington, DC: AERA.

National Research Council. (2014). *Developing Assessments for the Next Generation Science Standards*. Washington, DC: The National Academies Press. https://doi.org/10.17226/18409.

Patton, M. Q. (2015). *Qualitative research & evaluation methods: Integrating theory and Practice* (4th ed.). Thousand Oaks, CA: Sage Publications.

Shepard, L., Hammerness, K., Darling-Hammond, L., & Rust, F., Baratz-Snowden, J., Gordon, E., Gutierrez, C., & Pacheco, A. (2005). *Assessment*. In L. Darling-Hammond & J. Bransford (Eds.), *Preparing teachers for a changing world: What teachers should learn and be able to do* (pp. 275–326). San Francisco, CA: Jossey-Bass.

Smarter Balanced Assessment Consortium (2019), *Interim Assessments Interpretive Guide*, https://portal.smarterbalanced.org/library/en/reporting-system-interpretive-guide.pdf .

Weiss, R. (1994). *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. New York, New York: The Free Press.

Wilson, M., & Draney, K. (2002). *A technique for setting standards and maintaining them over time*. In S. Nishisato, Y. Baba, H. Bozdogan, & K. Kanefugi (Eds.), Measurement and multivariate analysis (Proceedings of the International Conference on Measurement and Multivariate Analysis, Banff, Canada, May 12-14, 2000), pp 325-332. Tokyo: Springer-Verlag.

Yarbrough, D. B., Shulha, L. M., Hopson, R. K., & Caruthers, F. A. (2011). *The program evaluation standards: A guide for evaluators and evaluation users* (3rd ed.). Thousand Oaks, CA: Sage.

Glossary of Acronyms

Acronym	Gloss
AERA	American Educational Research Association
CAA	California Alternate Assessment
CAASPP	California Assessment of Student Performance and Progress
CA NGSS	NGSS for California Public Schools, Kindergarten through Grade Twelve
CAST	California Science Test
CCC	Crosscutting Concept (CA NGSS)
CCSS	Common Core State Standards
CDE	California Department of Education
CERS	California Education Reporting System
CSA	California Spanish Assessment
DCI	Disciplinary Core Idea (CA NGSS)
DIBELS	Dynamic Indicators of Basic Early Literacy Skills
DL	Digital Library
DOK	Depth of knowledge
EC	California Education Code
EL	English learner (student)
ELA	English language arts/literacy
ELPAC	English Language Proficiency Assessments for California
ESSA	Every Student Succeeds Act
ETS	Educational Testing Service
FIAB	Focused Interim Assessment Block
FKSAs	Focal Knowledge, Skills, and Abilities

Acronym	Gloss
GVC	Guaranteed Viable Curriculum
IAB	Interim Assessment Block
IAVS	Interim Assessment Viewing System
ICA	Interim Comprehensive Assessment
LEA	Local educational agency
MTSS	Multi-tiered system of support
NCME	National Council on Measurement in Education
NGSS	Next Generation Science Standards
ORS	Online Reporting System
PBIS	Positive behavioral interventions and support
PE	Performance Expectation (CA NGSS)
PLC	Professional Learning Community
PT	Performance task
RTI	Response to intervention
SBE	State Board of Education
SEP	Science and Engineering Practice
SE	Socioeconomically
SMART	Specific, measurable, achievable, relevant, and time-based
SPSA	School plan for student achievement
SWD	Student with Disabilities
TAG	CAASPP Technical Advisory Group

Detailed Descriptions of Figures with Image

Figure 2.1 Screenshot of the home page of the CAASPP web page (Page 2-12)

- Screen shot of CAASPP website home page. Navigation menu at top of page lists eight main topics: Home, About, Test Administration, Resources, Training, Get Involved, Calendar, and System Status.

- The Home page is activated, with eight buttons displayed:

  - Test Operations Management System (TOMS)
  - Test Administrator Interface for All Online Tests
  - Practice & Training Tests
  - Tools for Teachers
  - California Educator Reporting System (CERS)
  - Completion Status/Roster Management
  - Smarter Balanced Content Explorer
  - Smarter Balanced Interim Assessments

Figure 2.2 Interim Assessment Administration Resources in the CAASPP website. (Page 2-14)

- Screen shot of Smarter Balanced Interim Assessment resources available under the Resources topic in the CAASPP website.

- The text “These resources support the Smarter Balanced Interim Assessments” is followed by eight buttons. A brief description describes the purpose of selecting each button.

  - Interim Assessment Viewing System: Select this button to access the interim assessments for professional development and/or training purposes.
  - Test Operations Management System (TOMS): Select this button to assign user roles for Tools for Teachers and the California Educator Reporting System, and to view student test settings, including accommodations, before interim testing begins. Note: To create/manage student groups, go to the California Educator Reporting System.
  - Test Administrator Interface for All Online Tests: Select this button to access the Test Administrator Interface that is used to access all CAASPP online assessments including the summative, interim, and alternate assessments.
  - Completion Status/Roster Management: Select this button to access the system that will allow you to see the completion status for students taking the interim assessments.

- Hand Scoring Training Guides and Exemplars: Select this button to access the interim assessment hand scoring training guides and exemplars. Upon selecting this button, select the [Resources] tab at the top.
- Interim Assessment Hand Scoring System: Select this button to access the system that will allow you to score student responses to interim assessment items that require hand scoring.
- California Educator Reporting System (CERS): Select this button to access interim assessment results or, for group administrators only, create/manage student groups.
- Reporting System Sandbox: Select this button to access the sandbox training tool. Username and password are not required, but users are prompted to select a role before entering the sandbox.