2020 No. 091

California Assessment of Student Performance and Progress (CAASPP) Independent Evaluation: Comprehensive Final Report 2018–2020

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	Date:	November 20, 2020

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

This page was intentionally left blank.

Comprehensive Final Report 2018–2020

Table of Contents

Table of Contents (cont.)

List of Tables

Table of Contents (cont.)

List of Figures

This page was intentionally left blank.

Executive Summary

Pursuant to California *Education Code* (*EC*) Section 60649, the Human Resources Research Organization (HumRRO) conducted an independent evaluation of the California Assessment of Student Performance and Progress (CAASPP) from July 2018 through December 2020. The purpose of the evaluation was to provide objective technical advice and consultation to the California Department of Education (CDE) on activities supporting the implementation of the CAASPP System.

This Comprehensive Final Report includes a summary of the following three studies HumRRO conducted during the three-year independent evaluation:

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

The content of this report is a condensed version of the three annual and four stand-alone study reports HumRRO produced for this contract. The reports are publicly available on CDE's website and include complete details of all aspects of study design, data collection, and analysis, as well as conclusions and recommendations. Links to the evaluation reports are included in the Introduction to this report. The final chapter of this report contains our reflections on positive aspects of the evaluation and suggestions for improving future evaluation work.

In this Executive Summary, we present major findings and recommendations from our three studies. We conclude with a synthesis of our three-year evaluation regarding the quality and effectiveness of the CAASPP System components addressed by our studies.

Instruction and Student Learning Case Study

The primary goal of the two-year case study was to elicit concrete examples of how and why specific Smarter Balanced English language arts/literacy (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 Interim Comprehensive Assessments (ICAs); and the Digital Library (DL), which includes formative assessments and instructional resources.[1] Each year of the study, HumRRO collaborated with six to seven local educational agencies (LEAs), including one direct-funded charter, and a subset of their schools (34 schools across the two years) to study their use of the Smarter Balanced components.

[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.

Major Findings

Year one findings are presented in condensed form in chapter 2. Here, we present findings from year two of the study, when the most recent versions of Smarter Balanced components were in place. Differences in the components from 2018–2019 to 2019–2020 prevent direct comparisons of study outcomes each year. For example, because each year of the study was conducted with different iterations of the CAASPP System, there were differing numbers of IAs and corresponding Connections Playlists available to each year's study participants. Also, based on outcomes from year one, HumRRO made extra efforts in year two to recruit LEAs and schools that used IAs to inform instructional decisions beyond preparing for the summative assessments (i.e., the selection criteria for LEA participation differs between the two years). Finally, the COVID-19 school closures impacted access to and use of IAs and resulted in the statewide suspension of 2020 summative testing. Thus, year two LEAs had less time to administer IAs compared to year one LEAs, and year two LEAs did not administer the summative assessments.

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

- **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., ELA claims scores) 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 were given an opportunity to provide input about decisions around IABs, which allowed better alignment of assessments with their curriculum.

- **Digital Library.** The study schools reported extremely limited use of resources from the DL, 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 the availability of sufficient materials through their curriculum or other familiar sources as reasons for not using the DL.

Summary of Best Practices

Based on the full scope of first- and second-year findings across the studied LEAs, HumRRO identified a sample of best practices supporting effective use of CAASPP Smarter Balanced 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. The following four best practices were observed across both years of our study:

- 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 Smarter Balanced components while allowing grade-level or content-area PLCs flexibility regarding what 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.

We identified three additional best practices that were unique to year two of the study:

- 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.

Recommendations

After each year of the study, HumRRO reviewed the full scope of findings to develop suggestions for the CDE to consider as part of its continuous improvement of the CAASPP System. HumRRO identified recommendations and included them in each year's respective stand-alone report. Each stand-alone report also described how some recommendations were being addressed by planned changes and updates to the system. Some recommendations were the same or similar across both years of the study, and others were unique to one of the years. All recommendations made across both years of the Case Study are described in chapter 2, organized by topic (e.g., Training Opportunities and Online Resources), and accompanied by relevant updates made or planned to be made to the CAASPP System.

Here, we present recommendations based on findings from year two of the study, when the most recent versions of Smarter Balanced components were in place.

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 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

The CAST is a computer-based assessment administered to students in grades five, eight, and once in high school (i.e., grades ten, eleven, or twelve). 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 CA NGSS to allow the CAST to address the full breadth of the standards. Within the California Next Generation Science Standards (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 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.

Major Findings

This section provides a high-level summary of the findings from HumRRO's evaluation of the alignment between the CAST and the CA NGSS based on (a) evaluation of the test contractor's CAST development and design documentation and (b) analysis of CAST 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, most items were rated as multidimensional, and more than half the 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.

Recommendation

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

The CAA for Science is administered to eligible students in grades five, eight, and once in high school (i.e., grades ten, eleven, or twelve). 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.

Major Findings

This section provides a high-level summary of the findings from HumRRO's evaluation of the alignment between the CAA for Science and the Science Connectors (derived from the CA NGSS) based on (a) evaluation of the test contractor's CAST development and design documentation and (b) analysis of CAST item ratings by content experts.

- 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 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.

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; therefore, HumRRO offered no recommendations.

Summary

HumRRO's 2018–2020 independent evaluation found the specific CAASPP System components we studied were functioning reasonably well. Across the two-year period of performance of the Case Study, we made several suggestions for further improving the efficiency and efficacy of the Smarter Balanced components of the CAASPP System. These recommendations were driven by data we collected from educators in a small number of LEAs and schools that were selected based on their self-reported use of the various components. During the Case Study, the CDE, its vendors, and the Smarter Balanced Assessment Consortium developed and implemented significant improvements to support the use of Smarter Balanced components to inform instruction and improve student learning.

Results from our CAST and CAA for Science alignment studies were generally very positive and did not indicate that any major changes in test development or forms construction processes and procedures are needed. Our study of the CAST indicated generally good alignment of the test items to the CA NGSS, and we expect the CDE and its contractors will consider the single specific suggestion for further strengthening alignment and thus the validity of score interpretations. Our study of the CAA for Science indicated generally good alignment of the test items to the Science Connectors derived from the CA NGSS, with no recommendations for improvements.

We fully support the CDE's continued efforts to implement solutions to areas identified for improvement, internally and by our independent evaluation, as the CAASPP System matures.

Introduction

Organization of the Comprehensive Final Report

This report covers all activities performed during the contract period of 2018–2020. The primary activities were to conduct three research studies: the Case Study, a qualitative study that addressed use of Smarter Balanced assessments and resources, and the two science assessment alignment studies. Following are brief descriptions of the contents of each chapter in this report.

- Chapter 1, "Overview of the Evaluation," provides background information about the independent evaluation, describes development of the 2018–2020 Three-Year Evaluation Plan, and highlights general activities HumRRO conducted that were integral to the overall evaluation.

- Chapter 2, "Instruction and Student Learning Case Study," presents HumRRO's methods and data collection activities conducted during each year of the two-year Case Study. The goals of the study were to learn how educators use the CAASPP Smarter Balanced Assessment System components (i.e., summative and interim assessments and the Digital Library) to inform English Language Arts/literacy (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 local educational agencies (LEAs) and a small subset of each LEA's schools. A distinct set of LEAs and schools were studied each year. HumRRO conducted the following data collection activities for this study: (a) in-person focus groups/interviews, (b) monthly email polling with LEA and school educators, (c) end-of-year web-based focus groups with LEA and school points of contact, and (d) student questionnaires. The chapter provides, for each research question, the overarching themes and unique aspects discovered in the LEAs' use of Smarter Balanced components. The chapter concludes with best practices and recommendations for effective use of the Smarter Balanced components. This chapter is a condensed version of the two stand-alone Case Study reports. The year one report is publicly available on CDE's website (https://www.cde.ca.gov/ta/tg/ca/documents/caasppimpactcasestudy19.pdf) and the year two report will be available in early 2021.

- 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: (a) research questions to investigate the alignment of CAST to the California Next Generation Science Standards (CA NGSS), (b) a summary of the methods and data collection activities completed, (c) alignment acceptability criteria that HumRRO developed, (d) evaluation of CAST contractor documentation, and (e) final outcomes of analysis of the alignment data. The stand-alone CAST Alignment Study Report is publicly available on CDE's website (https://www.cde.ca.gov/ta/tg/ca/documents/castalignmentstudy0420.pdf).

- 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: (a) the study’s research questions to investigate the alignment of the CAA for Science to the Science Connectors and Focal Knowledge, Skills, and Abilities (FKSAs) derived from the CA NGSS; (b) a summary of the methods and data collection activities completed; (c) alignment acceptability criteria that HumRRO developed; (d) evaluation of CAA for Science contractor documentation; and (e) final outcomes of analysis of the alignment data and. The stand-alone CAA for Science Alignment Study Report is publicly available on CDE’s website (https://www.cde.ca.gov/ta/tg/ca/documents/caas19alignmentstudyrpt.pdf).

- Chapter 5, “Reflections on Evaluation Activities,” concludes the report with our reflections on methods followed and approaches taken to conduct the 2018–2020 independent evaluation of the CAASPP System. We also offer considerations for what worked well or might be improved for future evaluations.

Chapter 1: Overview of the Evaluation

Background

The California Assessment of Student Performance and Progress (CAASPP) System was established by California Assembly Bill (AB) 484 on October 2, 2013, and has been the statewide student testing program since January 2, 2014. A highlight of the CAASPP System is its shift in focus from accountability alone to a comprehensive plan for promoting high-quality teaching and learning for all students, including students with disabilities (SWDs) and English learners (ELs).

The legislation requires an independent evaluation to provide objective technical advice and consultation on activities supporting implementation of the CAASPP System. The evaluation is defined in California *Education Code* (*EC*) Section 60649, which states that evaluation "activities may include, but not necessarily be limited to, 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 California Department of Education (CDE) awarded the Human Resources Research Organization (HumRRO) an initial contract to conduct an independent evaluation of the CAASPP System beginning July 2015 and concluding December 2017. During that contract, HumRRO conducted five research studies and produced three annual reports (2015, 2016, and 2017) and a Comprehensive Final Report of independent evaluation activities, findings, and recommendations. The CDE awarded HumRRO a second contract to continue its independent evaluation of the CAASPP System beginning July 2018 and concluding December 2020. The second evaluation contract called for three studies and three annual reports (2018, 2019, and 2020). HumRRO's annual reports include all data analyses pursuant to *EC* Section 60649. Additionally, HumRRO produced stand-alone reports that include full details of all aspects of each study conducted during 2018–2020. HumRRO's reports are publicly available on the CDE's website, https://www.cde.ca.gov/ta/tg/ca/caaspprptstudies.asp.

This Comprehensive Final Report includes condensed versions of content originally presented in the 2018, 2019, or 2020 annual reports or in the stand-alone reports for each study.

Three-Year Evaluation Plan

California *EC* Section 60649 requires development of a plan to assess independent evaluation activities, and it prohibits duplication of studies conducted as part of a federal peer-review process or by CDE assessment contractors. The CDE specified in its Request for Proposals (RFP) that the 2018–2020 independent evaluation focus on the following CAASPP System components:

- Smarter Balanced Summative Assessments (ELA and mathematics), required for grades three through eight and grade eleven, comprised of a computer-adaptive test (CAT) and a performance task (PT).

- Smarter Balanced Interim Assessments (ELA and mathematics), optional assessments designed for grades three through eight and grade eleven, available for use by educators from kindergarten through grade twelve to monitor student performance throughout the school year.

- Smarter Balanced formative assessment measurement tools and resources, available in the Smarter Balanced Digital Library (DL).[2]

- California Science Test (CAST), aligned with the California Next Generation Science Standards (CA NGSS), required for students in grades five, eight, and once in high school.

- California Alternate Assessment for Science (CAA for Science), administered in grades five, eight, and once in high school to students with the most significant cognitive disabilities.

The CAASPP System's theory of action, along with CDE priorities and the timeline for administration of the assessments, guided the goals and schedule of three research studies, which together comprised 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 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 presents an overview of the primary activities and schedule for the research studies HumRRO conducted during the 2018–2020 evaluation.

Table 1.1 Primary Activities and Schedule for Each 2018–2020 Evaluation Study

Study Title	Primary Activities and Schedule
Instruction and Student Learning Case Study	• For each year of the two-year study, collaborated with and gathered extensive qualitative data from a small sample of schools and LEAs, purposefully selected based on their use of CAASPP components and resources. The small sample aimed 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).

[2] During 2018–2020, 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 September 2020.

Table 1.1 (cont.)

Study Title	Primary Activities and Schedule
Instruction and Student Learning Case Study *(cont.)*	• Investigated the context and various approaches used by the small sample of schools and LEAs to implement and integrate the Smarter Balanced Assessment System components to inform instruction and improve student learning. • Conducted data collection activities in 2018–2019 with initial set of LEAs and schools. Conducted second year of data collection activities in 2019–2020 with new set of LEAs and schools. • Completed year one data analysis in 2019 and completed year two data analysis in 2020. • Developed the first and second stand-alone Case Study reports. The reports describe in detail the 2018–2019 and 2019–2020 findings of the studied LEAs' and schools' use of CAASPP components for purposes of informing instruction and student learning.
CAST Alignment Study	• Conducted data collection activities in 2018–2019 to evaluate the degree of alignment between the CAST test items and test forms with the CA NGSS. • Completed data analysis in 2020. • Developed stand-alone CAST Alignment Study Report, providing validity evidence suitable for submission for federal peer review under the Every Student Succeeds Act (ESSA) and informing future item development.
CAA for Science Alignment Study	• Conducted data collection activities in 2019–2020 to 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. • Completed data analysis in 2020. • Developed stand-alone CAA for Science Alignment Study Report, providing validity evidence suitable for submission for federal peer review under the Every Student Succeeds Act (ESSA) and informing future item development.

Staying Updated on the CAASPP System

The CAASPP System has continued to evolve during the course of this independent evaluation. To collect timely information about planned changes to the CAASPP System that might impact our work, HumRRO researchers engaged in several key activities. First, HumRRO's Project Management team attended the semiannual meetings conducted by the testing contractor to discuss with CDE staff the test administration and development activities planned for the upcoming nine to twelve months. Second, HumRRO's Project Management team attended three CAASPP TAG meetings each year, during which assessment plans and research outcomes for the Smarter Balanced assessments as well as for the CAST and CAA for Science were discussed. HumRRO presented progress reports on our studies at TAG meetings to obtain feedback from TAG members, testing contractor staff, and CDE staff that informed whether adjustments were needed to align our studies with the current CAASPP System implementation. Additionally, HumRRO's Project Management team conducted biweekly teleconference calls with the CDE contract monitor and other CDE staff, such as the Science Program staff and the DL and Interim Assessment (IA) Liaison, to proactively troubleshoot any challenges that might arise, such as changing CDE priorities or revised timelines for assessment development. When all CAASPP testing was suspended in March 2020 due to the coronavirus disease (COVID-19) outbreak, the biweekly meetings helped to develop solutions to continue our Case Study research with collaborating LEAs and schools that were transitioning to distance learning.

HumRRO conducted additional background research critical to maintaining knowledge of current Smarter Balanced components and processes. Several researchers observed annually at least two educator workshops supported by the CDE, such as the *2019 Summer Institute — Analyzing Student Work and Using the Interim and Digital Library Systems to Inform Teaching and Learning* and the *2019 California Assessment Conference*. Information from these workshops provided context for the Case Study's design and execution, as well as interpretation of results. Researchers also reviewed new publicly available online information to understand how updates to CAASPP components were presented to California teachers, administrators, and district staff. Finally, HumRRO researchers subscribed to the Assessment Spotlight, CDE's weekly email that provided information about CAASPP to educators from kindergarten to grade twelve.

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

Safeguarding Confidential Data

While carrying out our 2018–2020 Independent Evaluation Plan, HumRRO adhered to policies that protect and monitor access to sensitive information, such as student-level data. Cognizant of federal policies such as the Family Educational Rights and Privacy

Act (FERPA) as well as policies pertaining to governmental agencies in California and those specific to the CDE, our security program for the CAASPP evaluation focused on three key areas:

- Proper administration of non-disclosure agreements and implementation of the "need-to-know" principle for all personnel working on the contract.
- Comprehensive security training on specific security requirements related to our CAASPP work, including but not limited to specific data security and incident report procedures.
- Clear explanation of pertinent laws, regulations, and procedures related to safeguarding certain types of information relevant to the contract.

Taken together, these areas of our security program ensured all security procedures were administered in an efficient and effective manner.

This page was intentionally left blank.

Chapter 2: Instruction and Student Learning Case Study

Overview

The two-year 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 components of the CAASPP System (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. During the 2019–2020 school year, the second year of the study, HumRRO collaborated with six LEAs, including one direct-funded charter school. This chapter summarizes the activities and findings across both years of the study. The year one report is publicly available on CDE's website (https://www.cde.ca.gov/ta/tg/ca/documents/caasppimpactcasestudy19.pdf) and the year two report will be available in early 2021.

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 the year prior to its participation (2017–2018 for year one LEAs and 2018–2019 for year two LEAs) and planned to continue implementation during the study year, 2018–2019 or 2019–2020, respectively (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 methods of data collection, as described further in this chapter, to provide an in-depth look at the implementation of CAASPP components 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 year two 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 distance learning. Similarly, one LEA and two of its schools agreed to participate in a follow-up focus group in October 2020, as they dealt with distance learning. HumRRO researchers express their 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, to provide context for the reported findings. 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 next section presents the data collection activities, including modifications made during implementation of the study, and approach to data analysis.

The final sections of this chapter present general findings regarding CAASPP component use across all the LEAs studied during year one and more in-depth findings organized by the research questions of the study for year two. This section includes HumRRO's evaluation of contextual implications, common experiences, best practices, and challenges. The chapter concludes with a summary of best practices observed across both years of the study, recommendations identified each year for further improvements to be made, and brief descriptions of recent and planned changes to the CAASPP System that address the recommendations. The outcomes of the Case Study 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.

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, which represents what was available for the participants in our year two study. Data collected from our year two participants occurred following many CAASPP System updates compared to what was available when we collected data from our year one participants.

All of 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 (Resources 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 audiences (e.g., classroom teacher, CAASPP coordinator), and provided links to archived videos and webcasts of sessions and materials.

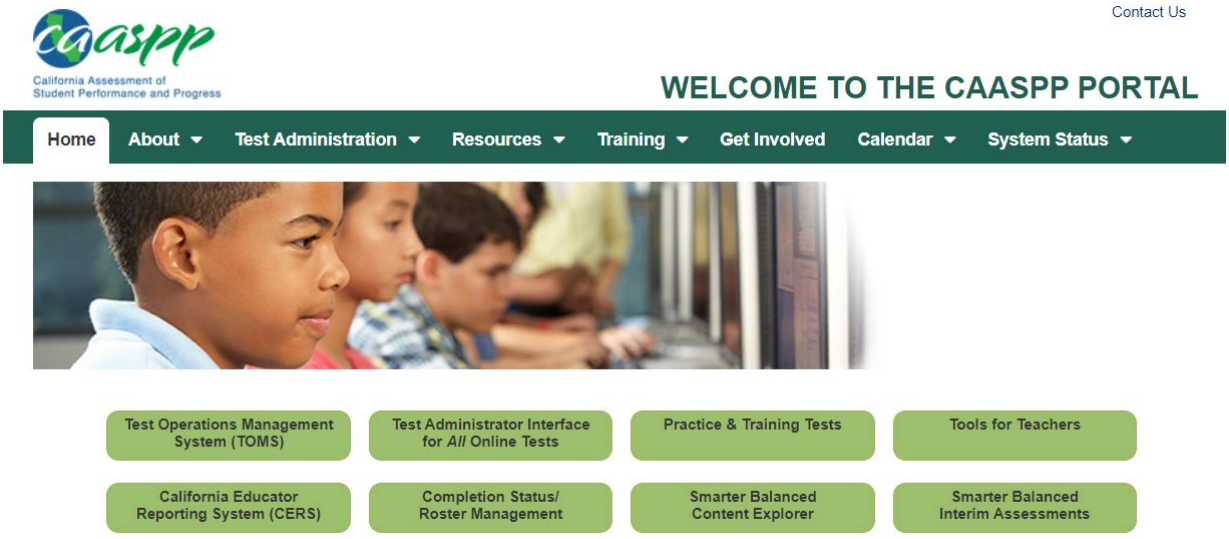

Figure 2.1 Screenshot of the home page of the CAASPP website.

CDE's online resources and in-person workshops for the CAASPP 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.

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.[3] 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, parents, educators, researchers). Individual student reports are available only to LEA CAASPP coordinators and school test site coordinators and to parents or guardians and are obtainable 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)

[3] 2019–2020 was not a typical year. Due to COVID-19 school closures, the Federal Government approved California's request for a waiver for 2019–2020 accountability testing.

Smarter Balanced Interim Assessments

The Interim Assessments (IAs) are not required but are available to California schools throughout the school year. Two main types of IAs in ELA and mathematics were offered during the 2018–2019 and 2019–2020 school years, 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; 42 new FIABs were made available for the 2019–2020 school year.

- ICAs cover the full range of targets, are built using the same blueprints as the summative assessments, and provide results on the same scale. In 2019–2020 ICAs were added to assess students in the ninth and tenth grades. These ICAs are similar to the eleventh grade IAB with grade-specific cut scores 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 secure online California Educator Reporting System (CERS) replaced the function of the former Interim Assessment Reporting System and became the vehicle for educators to view results for IABs and ICAs. 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. Clicking on 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. 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 into the content and format of educator training, which was attended by some Case Study participants, as well as an opportunity to learn about the latest system updates.

Smarter Balanced Digital Library

The Digital Library (DL) provided instructional resources for educators to use during daily instruction in support of the formative assessment process. Individual resources could 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 could 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 allowed teachers to be connected directly to DL resources that target their students' needs. The DL also provided 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

Research Questions

The Case Study addresses 13 key research questions pertaining to the CAASPP Smarter Balanced components of interest. Questions are organized into three general areas: (a) contextual questions and those pertaining to the full suite of CAASPP Smarter Balanced components, the Summative Assessments, IAs, and DL of formative 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 the research 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.

Table 2.1 CAASPP Smarter Balanced Components and Case Study Research Questions

Smarter Balanced Components Addressed	Research Questions for Sampled LEAs and Schools
All Components	1. What are the characteristics and contexts of sampled schools/LEAs that have implemented the full suite of Smarter Balanced components?
All Components	2. How does implementation of Smarter Balanced components vary across schools/LEAs? What instructions and supports are provided to educators for implementing the components?
All Components	3. What aspects of Smarter Balanced components are perceived as most beneficial for improving classroom instruction and student learning across schools/LEAs?
All Components	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?

Table 2.1 (cont.)

Smarter Balanced Components Addressed	Research Questions for Sampled LEAs and Schools
All Components	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?
All Components	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?
IAs 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?
IAs 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?
IAs 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?
IAs 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?
IAs 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?
IAs and DL	13. How is the DL used to improve classroom instruction?

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.

LEA Sample

For the Case Study, HumRRO's goal was to identify and recruit six LEAs in year one and six LEAs in year two (including one charter school each year) 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 due to the time commitment and competing priorities.

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 500 or more IABs in ELA and 500 or more 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 when identifying our year two sample. We learned that the

login data did not capture every use of the DL when resources were accessed indirectly (e.g., through IA reports). In addition, many logins turned out to be teachers who accessed the DL during professional development and never actually used the resources.

HumRRO used the DL and IAB criteria to prescreen potential LEAs for year one of the study and used the updated IAB criteria to prescreen potential LEAs for year two of the study. Each year, HumRRO administered an Eligibility Survey to all LEAs that met the minimum requirements. The survey for the first year was administered in early fall 2018, and the second year survey was administered in late summer 2019. 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 endorse the Case Study, invite LEAs to participate in the Eligibility Survey, and encourage their LEA's response. The 2018 and 2019 Eligibility Surveys were similar in content, though we made slight modifications to the 2019 version to ask if the LEA used IABs to inform classroom instruction beyond preparing for the summative assessments. 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 LEA CAASPP coordinators an invitation to complete the online survey. Table 2.2 summarizes survey respondents by LEA type for the first year of the study (overall 25% response rate) and interest in participating in the study, and table 2.3 summarizes the survey responses for the second year of the study (overall 33% response rate) and interest in participating.

Table 2.2 2018 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)	349	86	33	9	44
Charter	36	12	8	3	1
Total	385	98	41	12	45

Explanation of table contents: *Line 1 shows that we invited 349 non-charter LEAs to participate in our Eligibility Survey. Of these, 86 (or 25%) responded. Of the 86 respondents, 33 (38%) reported they would be potentially interested in participating in the Case Study, 9 (10%) reported they were possibly interested, and 44 (51%) 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 2018 and 2019 Eligibility Survey, HumRRO identified a list of the strongest

candidates to recruit for participation both years of the study. 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 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 in terms of demographics, but rather identified a sample of schools that were strong implementers of the CAASPP Smarter Balanced components.

Table 2.3 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.*

HumRRO encountered various challenges when recruiting and maintaining LEA participation during the two years of our study. During year one, one LEA dropped out mid-way through the study because of the time commitment and a replacement was identified for the final two months. During the year two recruitment period, multiple LEAs that met our recruitment criteria were dealing with nearby wildfires and associated power outages. Both years, some LEAs indicated participation in other studies or initiatives that would make it difficult to participate in the study. Despite these challenges, HumRRO identified seven LEAs in year one and six LEAs in year two for participation in the Case Study.

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 2018–2019 school year for year one LEAs, and for the 2019–2020 school year for the year two LEAs. 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 an honorarium for participating. Due to the COVID-19 school closures, HumRRO loosened requirements for the final months of the second year 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.

To preserve confidentiality and maintain anonymity, LEAs are identified only by number in this report (Y1-LEA-1 through Y1-LEA-7 for year one, and Y2-LEA-1 through Y2-LEA-6 for year two). Table 2.4 summarizes the characteristics of the seven participating LEAs in year one, and table 2.5 summarizes the characteristics of the six participating LEAs for year two, in terms of academic achievement in ELA and mathematics and select student demographics for the school year prior to their participation. Data in the tables are from 2017–2018 for the year one schools, and for 2018–2019 for the year two schools. The tables also indicate enrollment of students in the state or LEA who were in grades eligible for the CAASPP summative assessments those years.

Table 2.4 Characteristics (2017–2018 Data) of LEAs Participating in Year One of the Case Study

Case Study LEA #	Total Enroll-ment	# CAASPP Eligible Students	% Met or Exceeded ELA State Standards	% Met or Exceeded Math State Standards	% SE Dis-advant aged	% SWD	% EL
Y1-LEA-1	17,122	10,764	31%	23%	89%	8%	58%
Y1-LEA-2	4.270	2,188	59%	44%	49%	10%	26%
Y1-LEA-3	2,465	564	83%	73%	59%	9%	7%
Y1-LEA-4	4,882	2,656	30%	19%	91%	15%	37%
Y1-LEA-5	621,414	262,099	42%	32%	81%	14%	23%
Y1-LEA-6	3,926	1,976	45%	29%	47%	13%	3%
Y1-LEA-7	22,777	11,979	30%	20%	85%	12%	23%
All California	6,220,413	3,275,552	50%	39%	60%	12%	20%

Explanation of table contents: *Line 1 shows that the LEA we labeled Y1-LEA-1 had a total enrollment (across all schools, including those not participating in the study) of 17,122 students in 2017–2018. Of these, 10,764 were eligible to participate in the CAASPP summative assessments. Of those who took the summative assessment, 31% met or exceeded the ELA state standards, and 23% met or exceeded the math state standards. In Y1-LEA-1, 89% of students were socioeconomically (SE) disadvantaged, 8% were SWDs, and 58% were ELs.*

Table 2.5 Characteristics (2018–2019 Data) of LEAs Participating in Year Two of the Case Study

Case Study LEA #	Total Enroll-ment	# CAASPP Eligible Students	% Met or Exceeded ELA State Standards	% Met or Exceeded Math State Standards	% SE Dis-advant aged	% SWD	% EL
Y2-LEA-1	103,194	48,480	55%	46%	58%	14%	21%
Y2-LEA-2	13,870	7,051	38%	25%	84%	9%	31%
Y2-LEA-3	48,936	24,745	40%	27%	90%	12%	24%
Y2-LEA-4	32,138	17,015	81%	78%	6%	9%	5%
Y2-LEA-5	9,782	4,953	43%	28%	64%	13%	24%
Y2-LEA-6	1,833	1,093	69%	49%	25%	8%	8%
All California	6,186,278	3,189,956	51%	40%	61%	12%	19%

Explanation of table contents: *Line 1 shows that the LEA we labeled Y2-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 Y2-LEA-1, 58% of students were socioeconomically (SE) disadvantaged, 14% were SWDs, and 21% were ELs.*

As shown in tables 2.4 and 2.5, participation across the two years of our study includes a diverse range of school sizes, student achievement, and percentage of socioeconomically (SE) disadvantaged, students with disabilities (SWDs), and English learners (ELs). As shown, approximately half of the LEAs across the two study years have fewer than 50 percent of their students who met or exceeded the ELA state standards the year prior to their participation in the study. 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/).

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 collected data through teacher focus groups, POC and school principal interviews, and LEA POC interviews (and occasionally other staff).
- *Data from monthly polling of LEA and school POCs.* For five months of the study (December through April of each study year), HumRRO worked with POCs to gather LEA and school staff responses to one to three questions related to the use of Smarter Balanced components.
- *Data from end of school year web-based focus groups with LEA and school POCs*. HumRRO researchers conducted virtual focus groups with POCs toward the end of each study year. Each focus group was comprised of POCs across the study schools and grouped by type (e.g., LEA, elementary schools, middle schools, high schools).
- *Data from students*. In year one, HumRRO provided school POCs with instructions and protocols for conducting student focus groups and collecting written student responses through a questionnaire to gauge experiences with the summative assessments. The activity was modified and made optional in year two to account for distance learning and because the summative assessment requirement was waived; instead, HumRRO provided an online questionnaire and asked each school POC to consider collecting responses from their students to understand their experiences with one or more specific mathematics and/or ELA IABs.
- *Data from follow-up focus group with year one LEA (optional activity).* HumRRO invited year one LEAs and schools to participate in a fall 2020 virtual focus group. The purpose of the focus group was primarily to understand how CAASPP components were used in 2019–2020 and during distance learning. Only one LEA (Y1-LEA-7) and two of its schools chose to participate.

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 study 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 assigned 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 in year two, 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.

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 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).

Findings

This section summarizes the experiences of collaborating LEAs and schools. First, we provide a summary of year one findings by Smarter Balanced component (i.e., summative assessments, IAs, and DL). Next, we present in-depth year two findings by the 13 Case Study research questions. In reporting findings, we focus most predominantly on year two because the most recent versions of Smarter Balanced components were in place during that year of the study.

We present year one and year two findings separately to account for differences that prevent direct comparisons. For example, because the studies were conducted in separate years with different iterations of the CAASPP System, there were differing numbers of IAs and corresponding Connections Playlists available to each year's study participants. Also, based on outcomes from year one, HumRRO made extra efforts in year two to identify LEAs and schools that used IAs to inform instructional decisions beyond preparing for the summative assessments (i.e., the selection criteria for LEA participation differs between the two years). Finally, the COVID-19 school closures during year two impacted access to and use of IAs and resulted in the statewide suspension of 2020 summative testing. Thus, year two LEAs had less time to administer IAs compared to year one LEAs, and year two LEAs did not administer the summative assessment.

Year One Findings

The year one findings presented here were primarily identified through year one data collection activities. The study design called for follow-up focus groups in the fall of 2020 to obtain updates about the year one LEAs' use of CAASPP components during 2019–2020. These focus groups would have provided information about whether and how usage may have changed between the two study years following the various CAASPP System updates. Though we invited all year one LEAs to participate and offered an honorarium, due to the COVID-19 pandemic only one LEA and two of its study schools participated in a focus group in 2020. We incorporated information learned through this focus group with the year one findings.

Summative Assessments

School staff participating in the study reviewed summative assessment data from the prior year (2017–2018) during the first semester of the 2018–2019 school year. Some schools reviewed data as a school-wide team early in the year, while other schools did not do so until November or December. Delays in review of data were due to decisions made at the district level or confusion about the allowable uses of preliminary results. However, when scores were made available to districts in the Online Reporting System (ORS) in June 2018, the Deputy Superintendent of Public Instruction's letter to LEAs specifically stated the ORS results were not embargoed and encouraged use of the early results to inform educational programs and support local planning around the improvement of teaching and learning. The degree to which data were reviewed and used varied among schools. 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. Many also reviewed results by claim, and a few accessed target reports. Some teachers in our study had trouble recalling anything about the prior year's summative assessment scores, so they 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 to administer during 2018–2019.

The importance of summative assessment results was further stressed by the year one participants of an October 2020 follow-up focus group. These individuals expressed that the lack of spring 2020 summative assessment data was unfortunate for various reasons, including the lack of data to aid in student program placement prior to collection of classroom assessment data.

Interim Assessments

IAs were used by all schools in the study except one high school. 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.

Many teachers felt IAs were beneficial for preparing their students for the content, rigor, item types, and technology they would face on the summative assessments. Teachers believed exposure via IAs would benefit students' ability to demonstrate their knowledge and skills on the summative assessments. Some teachers saw additional benefits of IAs, finding them useful to identify gaps in student understanding and determine what content needed to be retaught. Some teachers chose to give the same IAB twice, as pre-test and post-test, to measure growth in student knowledge, though the CDE advises LEAs to be cautious in interpreting the results when IAs are used in this manner. Other teachers gave IAs only because of LEA-level mandates. There were mixed feelings on the utility of the IA Reporting System. In some cases, especially when results were accessed through the LEA's student information system, it seems teachers were not aware of the various features (e.g., reporting levels, item analysis, etc.) available to them through the IA Reporting System.

The LEA participant of the October 2020 focus group indicated they encouraged the nonstandardized administration of IAs at the onset of the 2019–2020 school year, and an administrator from one of the two participating elementary schools indicated they followed this guidance.

Digital Library

The study schools reported extremely limited use of DL resources. Most teachers were aware of the resources and had logged directly into the DL at least once; however, teachers noted time constraints, accessibility of sufficient materials through their curriculum or other familiar sources, and difficulty navigating through the system as reasons for not using the DL. Only two teachers across the entire study indicated the DL was beneficial for classroom instruction in 2018–2019. HumRRO did not investigate teachers' use of the alternate route to the DL resources via the IA Reporting System.

Year Two Findings

The year two findings are organized by research question, with research questions grouped into three general areas: (a) contextual questions and those pertaining to the

full suite of Smarter Balanced components, the Summative Assessments, IAs, and DL of formative tools; (b) the question related only to the Smarter Balanced Summative Assessments; and (c) questions related to the Smarter Balanced IAs and DL resources.

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 CAASPP theory of action 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 2.5. We considered two of our five non-charter LEAs to be large, two medium, and one small. Three of the LEAs were in northern California, one in central California, and two in southern California. 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% mathematics), 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 Y2-LEA-4 only six percent of its students were classified as socioeconomically disadvantaged and only five percent as EL. In contrast, Y2-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, geographic location, and student population, we noted consistencies of learning context among them. For example, all the studied LEAs devoted time for PLCs. School leadership and teachers corroborated this information; they expressed having set aside 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 Smarter Balanced components (exact training differed by school). LEA and school leadership across the study were also similar in reports of how they used their data, including CAASPP Smarter Balanced 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 studied schools had good access to technology. Similarly, district and school leadership, and most teachers participating in the case study, 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 mathematics. The schools selected for the study had used IABs for multiple years, and teachers were generally very familiar with how to administer them as well as how to report and use results.

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 year two in their 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. 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 DL. Though the LEAs and school administrators did not require its use, they made sure teachers were aware of its availability. The majority of 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 (e.g., CAASPP Summer Institute). School administrators, CAASPP site leads, and often teachers were provided training (e.g., for administering IAs, accessing reports, and IA handscoring) by LEA staff.

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 year two study generally found IAs to be the most beneficial aspect of the CAASPP System 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, 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, school leaders acknowledged the importance of summative data for generating school-level goals and tracking achievement.

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 year two 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 preferred having access to administering IAs earlier in the academic year. Some noted IABs were not available in their LEA until September or October 2019. Similarly, some indicated it would be useful to have summative assessment results provided earlier in the year for planning purposes.

Despite the fact 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.

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 updated 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 when they would have found the information more useful upfront.

LEAs and schools were generally satisfied with recent technology changes to the CAASPP System. For example, some teachers 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, LEA staff recommended technology improvements regarding the student rostering required before administering IAs. LEAs across our study conducted rostering at their central office for all schools, and this process required many labor hours. LEA staff across the study would appreciate this process to be simplified.

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 made 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 assessments 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 used 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.

Teachers expressed some challenge in using summative assessment results to inform classroom decisions for various reasons including that some teachers received results for students who were no longer in their classrooms. Additionally, teachers in LEAs that mandated specific IABs and when to administer them found the required IABs did not always align with their curriculum.

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 in year two because students did not take the summative assessments in spring 2020. However, HumRRO invited school POCs to administer an optional online student questionnaire about IABs, which were widely used to help prepare students for the summative assessments. HumRRO collected data from students at four schools (two high schools, an elementary school, and a charter school for grades six through twelve) about their experiences with IABs during the 2019–2020 academic year.

Through these data, we learned that most students across classrooms participating in the questionnaire had experience taking IABs in a standardized manner. The most common reason students believed their teachers administered IABs was to see how well students learned various skills. The second most common reason was to practice certain skills, and third to find out what skills they have been taught or still need to learn. Approximately half the students recalled IAB results led to their teacher reteaching

skills, and just below 40 percent of the students noted the results were used to identify student gaps in knowledge.

Smarter Balanced Summative Assessments

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 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. Schools generally reported that they developed school-wide goals, and some also examined data at the subgroup level to identify if special attention to specific groups of students was required.

Interim Assessments and Digital Library

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?**

IAs were used by all schools included in the Case Study. Specifically, all schools used IABs and some used ICAs. Table 2.6 notes the number of schools that administered IABs in the state of California overall, and for each of our year two 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 mathematics 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 Year Two Case Study LEA, and by Subject Matter and Manner

Case Study LEA #	# 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 CA	5,713	1,095	477	618	692	403
Y2-LEA-1	74	488	230	257	325	163
Y2-LEA-2	22	1,697	733	963	1,500	196
Y2-LEA-3	50	1,142	422	720	617	526
Y2-LEA-4	35	1,356	618	738	851	505
Y2-LEA-5	13	795	142	653	491	304
Y2-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 mathematics 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 year two 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 ELA, Across Case Study Elementary Schools

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.

N/A indicates the IAB is not available at the grade level.

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 ELA, Across Case Study Middle Schools

Test Name	Grade 6 (N Schools=6)	Grade 7 (N Schools=5)	Grade 8 (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.

**Indicates Focused IAB.

N/A indicates the IAB is not available at the grade level.

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 ELA, Across Case Study High Schools

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.

N/A indicates the IAB is not available at the grade level.

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, in our year two study. 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 Case Study Elementary Schools

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

Table 2.10 (cont.)

Test Name	Grade 3 (N Schools=6)	Grade 4 (N Schools=6)	Grade 5 (N Schools=8)	Totals
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.

N/A indicates the IAB is not available at the grade level.

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 grades 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 Case Study Middle Schools

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.

N/A indicates the IAB is not available at the grade level.

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 grades 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 Case Study High Schools

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.

N/A indicates the IAB is not available at the grade level.

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 year two LEAs. Y2-LEA-4 and Y2-LEA-6 did not administer ICAs, and Y2-LEA-3 included only two schools that administered, on average, 2 ICAs. Schools administering ICAs at Y2-LEA-1, Y2-LEA-2, and Y2-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 in 2019–2020 Per School, Statewide, and by Year Two Case Study LEA

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
Y2-LEA-1	5	132	60	72	127	6
Y2-LEA-2	4	201	60	141	189	13
Y2-LEA-3	2	2	0.5	1.5	1	1
Y2-LEA-4	0	N/A	N/A	N/A	N/A	N/A
Y2-LEA-5	1	162	110	52	132	30
Y2-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 mathematics 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. There were variations in how LEAs chose to mandate IAs—two LEAs required specific mathematics and ELA IABs, while one only required a minimum number of IABs to be administered per subject area, but allowed teacher groups within each school to select which IAs. Schools in our study tended to administer, or planned to administer,[4] more than the requirement, or recommendation, of their LEA. High schools in our study generally did not assess twelfth grade students with IAs.

[4] Though this was the plan in 2019–2020, many schools were unable to administer all planned IABs due to COVID-19 school closures.

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. Except for a small number of teachers from LEAs where IABs were mandated, teachers across LEAs 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 used 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. 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. Teachers across schools reported using IA results to identify areas of weakness in ELA and mathematics and adjusting instruction accordingly.

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. However, 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.

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)?**

Most educators in the 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. There were, however, a small number of teachers who indicated use of specific resources including answer keys for writing tasks and mathematics performance tasks.

Best Practices

Based on the full scope of first- and second-year findings across the studied LEAs, HumRRO identified a sample of best practices supporting effective use of CAASPP Smarter Balanced 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. The following four best practices were observed across both years of our study:

- 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 Smarter Balanced components while allowing grade-level or content-area PLCs flexibility regarding what 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.

We identified three additional best practices that were unique to year two of the study:

- 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.

Recommendations

HumRRO reviewed the full scope of Case Study findings each year 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.

HumRRO identified recommendations following each year of the study and included them in each year's respective stand-alone report. Each stand-alone report also described how some recommendations were being addressed by planned changes and updates to the system. Some recommendations were the same or similar across both years of the study, and others were unique to one of the years.

In this chapter, we present all of the recommendations made across both years, beginning with those for which there were similarities across years by topic. This means the order of recommendations here differ from the order in which they appeared in the stand-alone reports. We include descriptions of CAASPP System updates that occurred during 2019–2020 (following year one) and those planned to be implemented during 2020–2021 or beyond (following year two). Some of the planned changes for 2020–2021 include re-envisioned professional development opportunities to allow for online delivery given the COVID-19 circumstances.

Training Opportunities and Online Resources

Across both years of our study, we identified recommendations related to training opportunities and online CAASPP resources. This section describes related recommendations across both study years.

Year One Recommendation: Continue providing regional training opportunities and updated online resources for LEA- and school-level staff. The in-person trainings and CAASPP.org and CDE 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. [2019 Report Recommendation 1]

CAASPP System Improvements that occurred after the 2018–2019 school year:

- The CDE hosted a statewide California Assessment Conference in October 2019. The three-day conference offered a variety of sessions for classroom educators to explore the connection between assessments and classroom instruction and to explore ways of using assessment resources for improved teaching and learning.

- Beginning September 3, 2019, educators began using a single username and password (i.e., single sign on) to access the various CAASPP and ELPAC online systems, including the Test Administrator Interface, Interim Assessment Systems

(Viewing System, Hand Scoring System, and Reporting System), Online Reporting System (ORS), DL, and Practice and Training Tests.

A second year one recommendation also focused on training opportunities:

Year One Recommendation: The CDE should encourage LEA and school leaders to provide local training opportunities, including time and resources, to help teachers (a) accurately interpret Smarter Balanced summative and interim assessment results, (b) implement existing and new Smarter Balanced components, and (c) learn about enhancements to Smarter Balanced components. LEA and school leadership receive CAASPP training on Smarter Balanced components, and sometimes these trainings are made available to teachers. However, many schools have not had the time to pass along information to all their staff. Some teachers had not tried logging on to the IA Reporting System since the many enhancements to it were launched in 2018–19. Most teachers in the small study sample had not explored the DL, often because they found logging on confusing or because they felt they had sufficient resources already. Teachers in the study who tried the DL noted frequently that navigation was difficult and time consuming, though some of these teachers may have been referring to earlier versions of the system before it was enhanced. [2019 Report Recommendation 4]

Our year two study findings indicated CAASPP System updates implemented after the 2018–2019 school year benefitted LEA and school staff during 2019–2020. However, year two data collection activities also supported a recommendation that the CDE continue to provide training opportunities and update online resources.

Year Two Recommendation: 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 CAASPP components, and (c) learn about enhancements to existing components. [2020 Report Recommendation 1]

Recent or Planned CAASPP System Changes:

- The CDE modified the previously held in-person Summer Institute to be a virtual Interim and Formative Assessment Training Series in October 2020. The training content was organized into learning modules and is 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 include assessment literacy, interim assessment resources and systems, hand scoring practice on interim assessments, and formative assessment processes using Tools for Teachers. Additionally, three live webinars provide additional guidance and support to local facilitators.

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

- The CDE will offer virtual mathematics 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.

An additional year two recommendation was generated related to CAASPP online resources:

Year Two Recommendation: 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. [2020 Report Recommendation 3]

Recent or 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.

Increase the Number of IAs

LEA and school staff participating in years one and two of the Case Study tended to value the IAs and requested more of them be developed and made available. We identified related recommendations discussed below.

Year One Recommendation: Regarding interim assessments, the CDE should work with the Smarter Balanced Assessment Consortium to provide an expanded pool of ELA and mathematics tests, including multiple versions of existing IABs, ICAs for grades nine and ten, and shorter interim assessments that examine student achievement at the target level. Teachers using the existing interim assessments find them of high quality and requested more options for tests for classroom use. [2019 Report Recommendation 2]

CAASPP System Improvements that occurred after the 2018–2019 school year:

- New Smarter Balanced ICAs in ELA and mathematics were available for administration to students in grades nine and ten in 2019–20, with different cut points for each grade level.

- New Smarter Balanced FIABs (approximately 40 were released) measure one to three targets compared to up to eight targets measured by the regular IABs. The FIABs measure smaller bundles of content to (a) give teachers a better understanding of students' knowledge and academic performance and (b) provide teachers with precise next steps for instruction.

Through year two of our study, we learned participants were pleased with the addition of FIABs and wanted to see more IABs and FIABs developed and available in the future. The following recommendation stemmed from year two findings:

Year Two Recommendation: 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. [2020 Report Recommendation 2]

Recent or Planned CAASPP System Changes:

- The Smarter Balanced Assessment Consortium plans to release approximately 90 more FIABs over the following two school years (2020–2021 and 2021–2022).

IA Hand Scoring

Hand scoring often came up as a challenge during year one data collection activities, resulting in the following recommendation:

Year One Recommendation: Regarding the hand scoring requirements of some interim assessments, the CDE should explore how to address concerns related to the challenges some LEAs and schools have finding time for training and hand scoring. Some teachers in our sample who participated in hand scoring found it an excellent professional development activity, and others found instructional value in reviewing scored responses. However, constraints on time and resources often caused schools to decide against giving IABs that involve hand scoring. Perhaps the CDE could include an option for scoring via artificial intelligence techniques (currently being explored by Educational Testing Service, ETS). At the local level, support could take the form of (a) increasing the number of in-person hand-scoring training opportunities, (b) expanding the number of participants in such training, (c) providing teacher release time to engage in hand-scoring activities, or (d) sharing examples of teachers enthusiastic about their experiences with hand scoring (e.g., the CAASPP in Action series). [2019 Report Recommendation 3]

Though our year two teachers also administered IABs that required hand scoring, they typically did not see it as a barrier to IAB use, and so HumRRO did not make a recommendation regarding hand scoring in the year two stand-alone report.

Rostering and Platform Usability

During both years of our study, we learned from LEA and school staff that rostering students within the IA reporting system was labor intensive. In year one we identified the following recommendation related to rostering for the CDE:

Year One Recommendation: Seek ways to streamline or provide additional guidance on rostering within the IA Reporting System, including recommendations regarding what access LEAs should be providing to their teachers. Some CAASPP coordinators found the CAASPP rostering process to be cumbersome, and for one LEA there was confusion in 2017–18 that resulted in teachers not having student-level results. In addition, some teachers would like more access than they are currently provided by their school or LEA. Accessibility of IA report features at the educator level is dependent upon the creation of rosters by the coordinator. Teachers may benefit if their CAASPP coordinators are given more direction regarding what level of access they should provide their teachers. [2019 Report Recommendation 5]

In year two, LEA CAASPP coordinators expressed some of the same concerns with the rostering system we noted in year one. We also learned of other usability issues and combined suggestions to the CDE into the following recommendation:

Year Two Recommendation: 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. [2020 Report Recommendation 5]

Recent or Planned CAASPP System Changes:

- The 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 proprietary 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.

Resources for Instruction and Student Learning

During year two of the study, we identified two recommendations focused on adding or improving resources to assist teachers and students with instruction and self-learning:

Year Two Recommendation: 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. [2020 Report Recommendation 4]

Year Two Recommendation: 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 a 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. [2020 Report Recommendation 6]

Recent or 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, and includes high-quality materials that were reviewed by the State Network of Educators,[5] and 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 within the formative assessment process strategies and accessibility strategies, and offers options and ideas for differentiation of and student access to content.

- The CDE hosted 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 all educators who register.

[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.

Chapter 3: California Science Test Alignment Study

Overview

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.

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 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 CAST scores. 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. A full stand-alone report details the alignment approach, methods, and results (Dickinson, Thacker, & Hardoin, 2020).

Study Design and Data Collection

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 the 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* (page number in parentheses), the following eleven standards were identified for review:

Standard 1.9 (p. 25). 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 (p. 26). 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 (p.26). 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 (p. 43). 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 (p. 64). 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 (p. 67). 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 (p. 85). 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 (p. 85). 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 (p. 87). 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 (p. 89). 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 (p. 196). 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.

CAST Alignment Workshop

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 used in this study are grounded in the Webb alignment method (1997, 1999, 2002). HumRRO drew from Webb's concepts of categorical concurrence, depth-of-knowledge consistency, range-of-knowledge correspondence, and balance-of-knowledge correspondence and the principles of Webb alignment 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. 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 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 (grades five and eight 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 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 independently rating the remaining operational items.

Item ratings were generated via the following steps:

1. Panelists independently reviewed test items 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 agreed on 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.

Findings

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 3.2 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 3.2 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, 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 more than one of the following: a DCI, CCC, and/or SEP (and are therefore 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.

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 SEP and CCC. 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, 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.

In the evaluation of the 2018–2019 operational administration, the item-person maps generally depict item difficulty being aligned with students' ability. For all three grades, the distribution of item difficulties generally aligns 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 multi-point 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, Achievement Level 1, 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.

This page was intentionally left blank.

Chapter 4: California Alternate Assessment for Science Alignment Study

Overview

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.

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 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 is 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, two medium, and one high complexity test items (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 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 practices in test validation that facilitates using alignment study results in an overall validity argument. A full stand-alone report details the alignment approach, methods, and results (Dickinson, Thacker, & Hardoin, 2020).

Study Design and Data Collection

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?

Documentation Review

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 (*page number in parentheses), the following eleven standards were identified for review:

- Standard 1.9 (p. 25). 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 (p. 26). 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 (p. 26). 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 (p. 43). 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 (p. 64). 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 (p. 67). 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 (p. 85). 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 (p. 85). 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 (p. 87). 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 (p. 89). 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 (p. 196). 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.

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 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.** 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 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) Science 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 (grades five and 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 reach 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 independently reviewed test items and assigned ratings of:
    a) Science Connector measured by item
    b) FKSAs or 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 Science Connector, FKSA, or EU and cognitive complexity for each item.
4. Panelists agreed on 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. 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.

Findings

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 specific 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 aligned to a Science Connector. Similarly, all performance tasks at all three grade levels measured multiple Science Connectors, EUs, and 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 grades five and 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: Reflections on Evaluation Activities

This chapter offers our reflections on selected approaches taken to conduct the 2018–2020 independent evaluation of the CAASPP System. Aside from the technical outcomes of our activities, succinctly presented in the Executive Summary and reported in more depth in the study chapters and stand-alone reports, we note here a few key examples of what worked well and suggest considerations for what might be improved for future evaluations. We will address the following two activities that supported our evaluation:

- Collaborating with local educational agencies (LEAs) and schools
- Working with the CAASPP testing contractor

Collaborating with LEAs and Schools

HumRRO successfully recruited 13 LEAs and 34 schools within those LEAs to collaborate in the data collection activities over the course of the two-year Case Study. Our experience indicates the following processes and features were important to building and maintaining effective collaborations with LEAs:

- *Engage the CDE and LEA leadership in the recruitment process.* The initial outreach message was sent to superintendents and LEA CAASPP coordinators of the targeted LEAs by the Director of CDE's Assessment Development and Administration Division. This communication endorsed the importance of the study, alerted LEA leadership to the opportunity, and encouraged participation. HumRRO's subsequent communications were directly with the LEA CAASPP coordinators, knowing their leadership had been informed of the evaluation goals.

- *Clearly state expectations for participation in data collection activities in a signed Memorandum of Understanding (MOU) with the LEA.* Each LEA's agreement to collaborate on the Case Study was intended to foster a working relationship in support of data collection. Our recruitment process included an in-depth conference call with each LEA CAASPP Coordinator and other district leaders to discuss expectations and schedules for each data collection activity. HumRRO provided interested LEAs with an MOU that clearly outlined the study's research questions and described the commitments required by the LEA and its participating schools.

- *Offer honoraria to encourage LEA- and school-level participation in data collection activities.* HumRRO's contract included funds for honoraria to be paid to LEAs and schools for the time commitment required to collaborate on the study. LEA and school volunteers served as points of contact who gathered documentation, participated in interviews, coordinated focus groups and monthly polling with teachers, responded to monthly polling questions, coordinated administration of student surveys, and communicated regularly with HumRRO

researchers. Though a very small number of LEAs and schools were unable to accept the honoraria due to district policies, the offer served as an incentive for most LEAs to agree to participate.

- *Review preliminary findings with LEA- and school-level collaborators.* During year two, HumRRO conducted web-based meetings with participants from the studied LEAs and schools to critique high-level summaries of key findings from interviews, focus groups, and monthly polling. These meetings provided an opportunity for LEA and school staff to confirm or help refine HumRRO researchers' interpretations of the data.

Based on challenges HumRRO experienced during the Case Study, we suggest future evaluators consider the following suggestions when enlisting similar voluntary support from LEAs and schools:

- *Recruit for participation in the studies as early as possible.* For both years of the study, the elapsed time from first contact with the first LEA to the receipt of the signed MOU from the final LEA was approximately three to four months. This was nearly double the anticipated length of time expected to achieve agreements from the LEAs to participate. We found these three challenges to the planned schedule to be typical: (a) limited availability of LEA staff to meet with HumRRO to learn more about the study, (b) time needed to gain support from and identify a point of contact for each of the recruited study schools in the LEA, and (c) time for the LEA point of contact to gain approval to join the study and return the signed MOU.

- *Engage a subset of LEAs to commit to the full period of the study.* The evaluation contract began in July 2018, and HumRRO developed and fielded the 2018 Eligibility Survey and recruited LEAs and schools for year one as soon as possible. It was very challenging to find participants willing to join the study who were using IABs extensively. For this reason we did not require a two-year commitment in the MOU. Ideally, at least some year one LEAs would have agreed to participate in year two of the study when asked, so changes in use of Smarter Balanced components, as well as new uses of enhancements to the system, could be observed over time. Reasons given by LEA CAASPP coordinators for declining to join year two seemed mostly due to theirs and their teachers' schedules becoming even more demanding due to the new science assessments as well as district initiatives. Had it been possible to begin recruitment sooner, several LEAs could have been selected for the two-year commitment and could have come on board soon enough for year one participation.

- *Promote full participation in data collection activities*. During each year of the Case Study, LEAs and schools had varying degrees of participation, despite having committed to all activities described in the MOU. In some cases the reduced level of school participation seemed to result from the school having been pressured to join the study by the LEA. In other cases, lack of full

responses (such as to monthly polling) seemed to be a simple fact of competing priorities for limited teacher time. Perhaps a greater dollar amount in the honorarium would provide the necessary support for more consistent release time through the year for teacher participation.

- *Allow flexibility when situations warrant it.* Communities in California experienced unexpected challenges throughout the Case Study period. This included the COVID-19 pandemic and historical wildfires. HumRRO adjusted our participation expectations according to these situations so we could continue to collect quality information while not unduly burdening already stressed schools and LEAs.

Working with the CDE and the CAASPP Contractor

HumRRO was generally very pleased with the support from the CDE CAASPP contractor, and we were able to work effectively with ETS staff to obtain a variety of types of data throughout the 2018–2020 evaluation. The following activities contributed to this successful coordination between HumRRO and ETS:

- *The CDE's Request for Proposals (RFP) specified an orientation meeting in the scope of work for the evaluation.* CDE's RFP for the evaluation called for an orientation meeting within two weeks of the contract start date. In July 2018, HumRRO presented the draft 2018–2020 Three-Year Evaluation Plan to key staff from the CDE and CAASPP contractor (ETS). This project planning and coordination meeting was conducted at CDE offices in Sacramento and was attended by ETS's Senior Director of Operations, Strategic Advisors, and Assessment Specialists; CDE's Director of Assessment Development and Administration, the administrator of the Psychometrics, Evaluation, and Data Office, the contract monitor for the evaluation, several assessment administrators and consultants, and the Fiscal Manager; and HumRRO's Project Management Team and notetaker. Participants reviewed and discussed the overall schedule and data needs for the three evaluation studies and clarified their roles and responsibilities. During the meeting, critical next steps for implementing the evaluation plan were identified, and a foundation for open and effective communication among all parties was built. After the meeting, HumRRO distributed meeting minutes with action items designated for the CDE, ETS, and HumRRO.

- *The CDE included the evaluator in at least one annual CDE planning meeting with the CDE CAASPP contractor each year during the period of performance.* HumRRO's Project Management Team observed ETS's semi-annual meetings to gain a detailed overview of all the CAASPP activities to be undertaken in the next six to nine months and consider how those activities might impact the planned evaluation activities. Learning firsthand about any concerns regarding schedules or processes contributed to the effectiveness of the evaluation and the evaluator's requests for data and other information. Additionally, awareness of improvements or changes being made to the CAASPP System prior to their implementation was essential to conducting the evaluation.

- *The CAASPP contractor designated a point person to support the evaluator on each of the research studies*. HumRRO's research studies involved the Smarter Balanced Assessment System, the California Science Test (CAST), and the California Alternate Assessment for Science (CAA for Science). Each study had its own schedule of key milestones and unique requirements for documentation, data, and support from the testing contractor. The identification of a primary point-of-contact for each study streamlined HumRRO's coordination with the CAASPP contractor and helped ensure a timely solution to challenges that arose during evaluation activities (e.g., the need to resolve discrepancies in test specifications and metadata for the alignment workshops).

In conclusion, 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.

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 2018–2019 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., & 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
DCI	Disciplinary Core Idea (CA NGSS)
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
IAB	Interim Assessment Block
IAVS	Interim Assessment Viewing System
ICA	Interim Comprehensive Assessment
LEA	Local educational agency
NCME	National Council on Measurement in Education
NGSS	Next Generation Science Standards
ORS	Online Reporting System

Acronym	Gloss
PE	Performance Expectation (CA NGSS)
PLC	Professional Learning Community
PT	Performance task
SBE	State Board of Education
SEP	Science and Engineering Practice
SE	Socioeconomically
SWD	Student with Disabilities
TAG	CAASPP Technical Advisory Group

This page was intentionally left blank.

Detailed Descriptions of Figures with Image

Figure 2.1 Screen shot of the home page of the CAASPP website (Page 19)

- 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 22)

- 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 tis 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.