

California Department of Education

Executive Office

SBE-003 (REV. 11/2017)

imab-adad-jul24item05

# California State Board of Education July 2024 Agenda Item #14

## Subject

Request to Ratify the California Department of Education Grant Application for the Competitive Grants for State Assessments Program.

## Type of Action

Action, Information

## Summary of the Issue(s)

On May 23, 2024, the California Department of Education (CDE) submitted a federal grant proposal for the Competitive Grants for State Assessments (CGSA) Program. The CDE requested $3,996,009 which would be expended over four years beginning October 1, 2024. If funded the grant would support the development, piloting, field-testing, and evaluation of a set of science performance tasks embedded in learning (PTELs). The proposed Innovative Measures: Performance Assessment for California’s Tomorrow (IMPACT) project aims to support science learning and instruction by providing teachers and students with authentic experiences that measure science through doing science. The proposal aligns to the prior actions and goals for science innovations by the California State Board of Education (SBE).

## Recommendation

The CDE recommends that the SBE ratify CDE’s application for the CGSA Program.

## Brief History of Key Issues

Information regarding this federal grant opportunity was made available in the Federal Register on March 8, 2024. Grant application requirements are available at <https://www.federalregister.gov/documents/2024/03/08/2024-04972/applications-for-new-awards-competitive-grants-for-state-assessments-program>.

The purpose of this grant is to support the design of statewide assessment systems that promote a deeper understanding of academic achievement among all student groups. The work must prioritize support for effective instruction and building educator capacity through the development of high-quality assessments of student learning and strategies that allow educators to use data from assessments to inform instruction. In addition, priority is given to projects that improve automated scoring of non-multiple-choice items and reduce the time to provide results.

In California’s CGSA Program application, the CDE proposes to carry out the work in the California Science Test (CAST) Concept Paper presented to the SBE in January 2024. If funded, the grant will support the development of a set of performance tasks embedded in learning (Common PTELs) for use during the school year in grade five, grade eight, and once in high school (grade ten, eleven, or twelve), as well as end-of-year performance tasks (Mini PTs). The CDE will also evaluate the performance tasks, develop scoring strategies that leverage innovative scoring to assess multiple dimensions of student performance, and provide professional development to educators.

If awarded, the project will utilize a rigorous iterative process to ensure the validity and reliability of the performance tasks. First, cognitive labs will be conducted with students from relevant grade levels, representing the diversity of student experience in California. Next, a pilot test will be conducted in a small group of schools that will inform revisions and improvements. Finally, a larger-scale field test of up to 20 schools and 3,000 students will be conducted to provide data for an evaluation of automated scoring. Demographic characteristics will be an important factor in selecting participants. A stratified sampling approach will be used to ensure that the participants are representative of California’s public-school population in terms of key demographic variables (e.g., socioeconomic status, rural/urban setting).

This grant was developed in collaboration with interest holders and experts. The CDE included letters of support from:

* California Next Generation Science Standards Collaborative with executive members Jane Steinkamp representing the Curriculum Instruction Steering Committee (CISC), Maria Chiara Simani representing the California Science Project (CSP), Jill Grace representing K-12 Alliance, and Michele Holcomb representing the California Association of Science Educators (CASE); and
* inquiryHub with Dr. William Penuel, Distinguished Professor at the Institute of Cognitive Science and School of Education at the University of Colorado Boulder.

Letters indicating interest in serving on the Advisory Committee were provided by assessment experts including:

* Dr. James Pellegrino, Professor Emeritus at the University of Illinois at Chicago (UIC) and a founding Co-director of UIC’s interdisciplinary Learning Sciences Research Institute
* Dr. William Penuel, Distinguished Professor of Learning Sciences and Human Development and Director of National Center for research in Policy and Practice, University of Colorado Boulder
* Dr. Richard Shavelson, Professor Emeritus, Stanford University Graduate School of Education
* Dr. Steve Schneider, Senior Director STEM Research and Entrepreneurship, WestEd
* Dr. Jill Wertheim, Director of SCALE Science, WestEd
* Dr. Aneesha Badrinarayan, Director of State Performance Assessment Initiatives, Learning Policy Institute
* Dr. Raymond Pecheone, Professor Emeritus and Executive Director of SCALE, Stanford University Graduate School of Education
* Dr. Tamara (TJ) Heck, Director of Science Assessment, CenterPoint Education Solutions

Letters of support for this proposal were submitted by the Association of California School Administrators, California Teachers Association, Los Angeles County of Education, and the California Stem Network. In addition, resumes of key CDE staff and expert partners in the field were submitted with the application.

The proposed project is described in Attachment 2.

## Summary of Previous State Board of Education Discussion and Action

In January 2024, the SBE approved the receipt of the CAST Innovations Concept Paper provided in Attachment 1, pursuant to Task 6.8 of the California Assessment System contract number CN220002. The CAST Innovations Concept Paper describes a possible system of assessments which includes Common PTELs and Mini PTs (<https://www.cde.ca.gov/be/ag/ag/yr24/documents/jan24item03.docx>).

## Fiscal Analysis (as appropriate)

The CDE has requested $3,996,009 in Federal grant funds to be used in the following ways over a four-year period beginning October 1, 2024, if awarded:

* $696,000 for CDE administrative costs (includes hiring one Education Programs Consultant) for salaries and benefits
* $3,300,009 for contractual costs with ETS

CGSA funding decisions from the U.S. Department of Education are expected to be announced in September 2024. If the grant is awarded, the CDE intends to use the grant funding in conjunction with the funding designated for assessment innovations as described in Task 6.8 of the California Assessment System contract number CN220002.

## Attachment(s)

* Attachment 1: May 16, 2024 Letter: California’s Application for the Competitive Grants for State Assessments Program (1 page)
* Attachment 2: Project Narrative (78 pages)
  + **Note:** The Project Narrative, one requirement of a much larger application packet, provides a comprehensive description of the proposed project. The full application is available upon request.

# Project Narrative

**Contents**

Introduction 2

Significance 5

Quality of the Project Design 11

Quality of Project Services 42

Adequacy of Resources 57

Management Plan 60

Project Evaluation 69

References 73

## Introduction

The California Department of Education (CDE) and State Board of Education (SBE) are applying for the Competitive Grants for State Assessments (CGSA) program targeting ***Absolute Priority 1*,** focusing on evaluating student deeper learning of science through the development of performance task assessments that emphasize the mastery of the California Next Generation Science Standards (CA NGSS) performance expectations in a knowledge-in-use education model (Pellegrino et al., 2018).

This initiative represents an effort to deepen science learning by creating a more comprehensive assessment system that extends beyond the current California Science Test (CAST). The current CAST blueprint was approved in 2017, and the test became fully operational in 2018–19 (CDE, 2017, 2022).

In its next iteration, the CDE and SBE seek to position the state’s science assessment system as a true partner in teaching and learning for all interest holders. We recognize that state assessments signal to the field the kind of teaching, learning, and performance that are valued, and that what is tested is closely related to what is taught. To fully realize the vision of the CA NGSS, the CDE and the SBE have begun to explore a more integrated approach to science instruction and assessment. Central to this approach are classroom performance tasks that develop science and engineering practices and scientific sensemaking in every grade level for formative and instructional use, along with performance tasks as part of the CAST—both in the classroom and on the End-of-Year assessment—that measure these abilities. The goal is to support science learning and instruction by providing teachers and students with authentic and engaging experiences that *measure* science through *doing* science, while also providing them with materials they can use in the classroom to better support authentic student engagement in sophisticated disciplinary work.

Specifically, the proposed project, titled **Innovative Measures: Performance Assessment for California’s Tomorrow (IMPACT)**, aims to develop and pilot the performance task components of a conceptual approach for an innovative science assessment system. The conceptual approach includes four components: (1) a bank of classroom performance tasks for teacher use, to be embedded in learning throughout kindergarten through grade twelve (K–12) (Classroom performance tasks embedded in learning [PTELs]); (2) Through-Year PTELs (Common PTELs) that are components of the summative assessment; (3) End-of-Year assessment in tested grade levels with a mini–performance task (Mini PT); and (4) opportunity-to-learn-in-science indicators (OTL Indicators) reflecting information about science learning opportunities in classrooms across the state. This CGSA proposal targets two of these components, the Common PTELs and the Mini PTs that are part of the End-of-Year assessment, which will provide opportunities for students to engage in scientific practices as part of the formal assessment system. Other components in the conceptual approach are being developed under separate auspices.

More specifically, we propose to develop, pilot, field-test, and evaluate a set of Common PTELs for use during the school year in grades five, eight, and once in high school, as well as the Mini PT section of the end-of-year CAST in those grade levels. The addition of the Common PTELs and the Mini PTs will fill a gap in the state’s current science test by highlighting the importance of teaching and assessing science practices, providing insight into the development of those practices and related science competencies for individuals and groups.

IMPACT also addresses ***Competitive Priority 2*,** focusing on improving assessment scoring to provide better and more timely information to inform instructional planning. The work proposed under this priority centers upon exploring both automated scoring and psychometric approaches to combining scores from the Common PTEL with those from the End-of-Year assessment with the Mini PT to produce valid indicators of individual and group standing on the CA NGSS. The proposed work will explore innovative automated scoring algorithms tailored to assess multiple dimensions in student performances in the performance tasks, including both conceptual understanding and engagement with science practices. Leveraging ETS artificial intelligence (AI) science-assessment research capabilities (Liu et al., in press), these algorithms will be used to evaluate the extent to which students’ constructed responses integrate science practices with the application of scientific knowledge, which can provide useful information for identifying strengths and improvement needs to support students’ scientific reasoning and sensemaking. For conceptual and methodological reasons, automatically analyzing students’ constructed responses and then combining scores from Common PTELs and End-of-Year assessments administered at different time points is not straightforward. Demonstration and empirical evaluation of practical approaches to automated scoring and information aggregation will help to advance the state of the art.

Overall, IMPACT aims to address the needs of California’s diverse student population, which includes a large proportion of English learner (EL) students. The project places a strong emphasis on active engagement in science practices that transcends language barriers to ensure that all students, including ELs, have equitable opportunities to demonstrate their scientific competency. Moreover, the project adopts an inclusive and asset-based design and scoring approach that values the multilingual and multicultural assets students bring to their scientific learning to promote equity and fairness by embracing the diversity of California’s student population.

## Significance

The proposed IMPACT project goes beyond the traditional assessment methods to ensure that every student has the opportunity to demonstrate their scientific proficiency. Through its innovative design approach and commitment to equity, IMPACT sets a precedent for transformative science education practices that prioritize student engagement in scientific practices essential for making sense of the world and fostering future success. The significance of IMPACT lies in its ability to address the need to measure students’ standing in relation to the CA NGSS more validly, build capacity among interest holders, improve services, expand access and equity, and enhance utility in varied settings.

**More Valid Measures of Next Generation Science Standards**

The need for more valid science measures has become increasingly evident since the release of the *Next Generation Science Standards*. States have been facing significant challenges related to designing high-quality assessments that measure the depth and breadth of the standards. The use of performance tasks within state science tests holds considerable promise for advancing science learning and teaching of the NGSS. Well-designed performance tasks can broaden the representation of the constructs comprising the standards by calling upon science practices not fully evoked in selected-response tests. In this way, such tasks can offer a more valid approach to measuring student understanding of scientific concepts and mastery of practices required by the standards. Because California’s adaptation of the standards preserves the essential three-dimensional structure of the NGSS, the proposed work has the potential to be applicable broadly to states that have also adopted versions of the NGSS.

The [*Next Generation Science Standards for California Public Schools, Kindergarten Through Grade Twelve*](https://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp)(CA NGSS) were adopted by the SBE in November 2013 (CDE, 2022). The CA NGSS emphasizes three interconnected and equally important dimensions (3D) of science learning:

1. Disciplinary Core Ideas (DCIs)
2. Crosscutting Concepts (CCCs)
3. Science and Engineering Practices (SEPs)

DCIs are ideas at the core of four disciplines of science: Earth and Space Sciences, Life Sciences, Physical Sciences, and Engineering. CCCs are concepts that apply across multiple disciplines of science and engineering (e.g., patterns, systems and system models). SEPs are real-world practices engaged by practicing scientists and engineers (e.g., asking questions, developing and using models, constructing explanations). Students are expected to demonstrate their understanding of science by integrating the three dimensions together to make sense of phenomena and address real-world problems. The CA NGSS intends to prepare informed, scientifically literate citizens with the skills to participate in an increasingly global and fundamentally technological society (NGSS Lead States, 2013). A solid K–12 science education equips students to seek opportunities in scientific fields and participate as scientifically literate citizens. Additionally, the CA NGSS envisions that science education should inspire future generations of scientifically literate students and support development of career and college readiness through collaboration among various interest holders, including states, local educational agencies (LEAs), schools, teachers, and students (CDE, 2023).

**Building Capacity**

IMPACT’s commitment to developing valid assessments not only ensures accurate measurement of student’s scientific proficiency but also serves as a foundation for building capacity among interest holders. Developing meaningful performance tasks will require co-design among educators, assessment specialists, community representatives, and policymakers. Such co-design will work to build capacity within the CDE and LEAs. By engaging interest holders in design, piloting, and field testing, the proposed project can enhance assessment and scientific literacy among interest holders.

Educators play a pivotal role in shaping assessment practices and addressing the needs of students. Community members (including parents) likewise contribute to setting conditions conducive to learning. By actively involving educators and community members in the co-design process, the project can ensure that the performance tasks align closely with instructional objectives and reflect the diverse needs of students across different grade bands. Additionally, educators and LEA leaders can provide valuable insights into the feasibility and practicality of implementing these assessments. Involving these various interest holders in the co-design and review of assessment development promotes a shared understanding of assessment goals, criteria, and expectations. This collaboration process builds trust and transparency, and improves the final result. Collaborative discussions around assessment-data interpretation and use can help deepen interest holders’ understanding of how assessment results inform instructional decisions, policy development, and resource allocation. The collaborative process will also foster a culture of continuous improvement, as iterative design promotes continued growth in assessment and scientific literacy among all those involved.

**Improving Services**

As interest holders collaborate in the co-design and review process, they develop a shared understanding of assessment goals, criteria, and expectations. This collaborative effort helps lay the groundwork for improving services by equipping interest holders to leverage assessment data to enhance instructional practices and allocating resources effectively. Using the performance tasks during the year can lead to improvements in educational services by providing educators with more timely and actionable information about student learning. Educators can gather data on student performance at regular intervals, allowing for ongoing monitoring of student progress and identification of trends. The timely nature of results from Through-Year performance tasks can enable educators to promptly adjust instruction to meet the evolving needs of their students. Additionally, Through-Year performance tasks facilitate data-driven decision-making at the classroom, school, and LEA levels. Educators can collaboratively analyze assessment data, identify patterns, and design and implement targeted interventions to support the diverse needs of students.

Integrating information from the Through-Year performance tasks with End-of-Year assessment should foster alignment between classroom instruction and assessment of the standards. This alignment benefits both students and educators by promoting familiarity with the three-dimensional nature of the standards and ensuring that assessments accurately measure what is taught in the classroom. By connecting Through-Year and End-of-Year assessment, students will gain a deeper understanding of learning objectives throughout the academic year. Consistency in assessment formats and content between the performance tasks at these two occasions reinforces the relevance of classroom instruction, helping students recognize the importance of integrating specific core concepts with science practices.

**Expanding Access and Equity**

California has one of the largest and most diverse K–12 student populations in America, with a significant portion being ELs or once ELs (39%). Eighty percent of students are of color, with more than half (about 56%) of the total being Latino/Latina. Among the rest, 20% are non-Latino/Latina White, 10% Asian, 5% African American, 2% Filipino, and about 7% other racial/ethnic groups or multiracial groups.[[1]](#footnote-2) Given this diversity in the California student population, it is crucial to ensure that assessment practices are accessible and equitable. The performance tasks developed in the proposed work will offer authentic and inclusive experiences. Following principles of universal design and socioculturally responsive assessment, the tasks will be student-centered, responsive to differences in prior knowledge, interests, ways of knowing, modality preferences, and lived experiences (Darling-Hammond, 1996, 2017). To ensure that EL students and others facing language challenges have equitable access to demonstrate their scientific understanding, opportunities to communicate in different modalities will be afforded. This inclusivity ensures that all students have the opportunity to engage meaningfully with scientific practices and demonstrate their learning in ways that resonate with their personal experiences. To further enhance equity, the project will strive for diverse representation on the development team to bring varied perspectives and experiences into the design process. Through these means, the proposed work will promote greater access and equity in science assessment.

**Enhancing Utility in Varied Settings**

Ensuring that all students have equitable opportunities to engage in rigorous science education sets the stage for maximizing the effectiveness of using assessment data across different contexts. A common set of principles will be used to design the performance tasks for this project. These design principles aim at promoting equity and maximizing utility across diverse educational settings. The principles prioritize authenticity, inclusivity, and responsiveness to students’ needs.

By adhering to these principles, the performance tasks in this project can serve as a model for performance task assessment development by educators at the local level. The co-design process will allow interest holders to customize performance tasks to align with their specific instructional goals and student demographics, which empowers educators to tailor performance tasks to meet the unique needs of their students. The collaborative approach to performance-task development will foster sharing of innovation, and best practices and applications, across various educational settings to address diverse student needs.

## Quality of the Project Design

Given the significance of IMPACT in addressing the critical needs of California’s diverse student population and promoting equitable access to high-quality science education, it is essential that the project design not only reflects the overarching goals but also outlines the systematic approach to achieving them. In this section, we first center the proposed work around the idea of performance tasks embedded in learning (PTELs). We next introduce a conceptual approach for an assessment system that capitalizes on that idea. Third, we describe a theory of action for that assessment system. We conclude by detailing the project goals, objectives, activities, timeline, and outcomes.

**Performance Tasks Embedded in Learning**

To assist teachers in the implementation of the CA NGSS and to support diverse students’ learning needs, the CDE and SBE have conceptualized an innovative assessment system focusing on performance tasks that can be embedded in learning and that can be included in briefer form on the End-of-Year assessment. A concept paper for this innovative assessment system was prepared by ETS and received with enthusiasm by the SBE in January 2024. The concept paper described an assessment system intended to achieve the following goals:

* **CA NGSS–Aligned**, emphasizing authentic use of the SEPs, together with the CCCs and DCIs, to make sense of phenomena and address real-world problems
* **Student-Centered**, highlighting the need to build tasks on students’ experiences, prior knowledge, and how they learn
* **Instructionally Relevant**, providing both models and information for instruction

To accomplish these goals, the performance tasks developed for use in the classroom and for the End-of-Year assessment must meet the following criteria:

* **Be authentic:** highlight and center the key concepts, modes of inquiry, and ways of learning in the discipline
* **Involve making sense of phenomena:** create opportunities to meaningfully engage with real-world phenomena by placing phenomena within contexts likely to be familiar to students
* **Represent the NGSS dimensions:**facilitate students’ application and integration of the DCIs, SEPs, and CCCs in service of sensemaking
* **Reflect and be responsive to learners:** follow principles of universal design and responsiveness to learners’ experiences and ways of knowing to ensure accessibility and opportunities for learners to show what they know and can do
* **Be useful for informing decisions that impact instruction:** help encourage, model, and inform high-quality curriculum and instruction

High-quality PTELs designed to reflect these features might aid teaching and learning if they

* provide authentic opportunities for students to engage in scientific sensemaking using expectations for conceptual understanding and practice established by the CA NGSS, and receive feedback or related instruction from teachers throughout the year that help students progress toward mastery;
* help teachers gain experience in understanding how to support NGSS-aligned teaching and learning, as well as how to assess evidence of student progress and proficiency; and
* are readily available in the form of a curated set of task resources to help teachers gain familiarity with their design, use, and scoring.

In these ways, PTELs have the potential to model and facilitate high-quality instructional opportunities for students to engage in science learning and for teachers to leverage effective science-teaching practices (see Badrinarayan & Darling-Hammond, in press). Furthermore, PTELs have the potential to play a role in the reimagining of the state summative assessment.

The potential for PTELs to positively impact instruction is evident based on prior research on performance assessment. Stecher (2010) reviewed the literature on several major state initiatives, including those of Vermont, Maryland, and Kentucky. He reported that performance assessment led to greater emphasis on problem-solving and communication in mathematics, and to more extended writing in language arts. More recently, one study done with New Hampshire’s Performance Assessment of Competency Education (PACE) system in K–12 English language arts and mathematics found statistically significant improvements in state achievement scores for participating versus nonparticipating students (Evans, 2019).

**Project Rationale**

The IMPACT project is an integral part of a larger conceptual approach aimed at innovating the CAST program centered around performance tasks, which is supported by multiple funding sources beyond the CGSA grant. The rationale for the IMPACT project stems from the broader initiative. The IMPACT project will be focusing specifically on the summative components of this conceptual approach, including both Through-Year and End-of-Year performance tasks. The first section of the project rationale includes a detailed introduction to the new conceptual approach, which offers a comprehensive picture of the key components. Following the conceptual approach, we describe the theory of action guiding the larger initiative. Contextualizing the IMPACT project within this larger initiative provides a clear theoretical foundation for the significance of the project and sets up the delineation of the goals and objectives elaborated in the subsequent section.

*A New Conceptual Approach for an Assessment System*

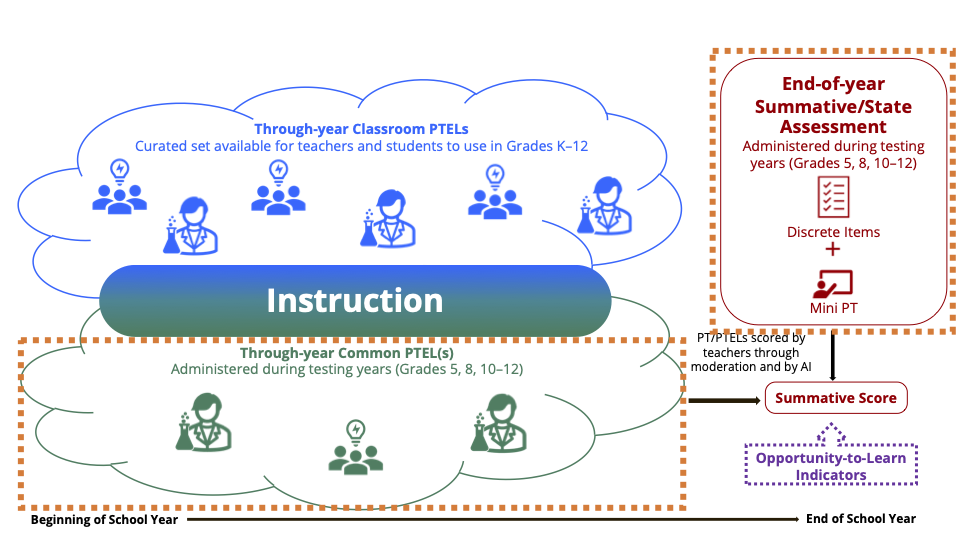
Through discussions with subject-matter experts, review of state and local performance assessment initiatives, and review of the research literature, we developed a conceptual approach that comprises four assessment-related system components. In this section, we describe all four components to provide insight about the broader vision guiding the proposed project. The components include the following:

* *Through-Year Classroom PTELs (Classroom PTELs),* readily available for teachers in all grade levels as tasks that support formative assessment and instruction of the CA NGSS. From a statewide bank, teachers would select tasks tagged to standards, grade levels, and curriculum contexts and use them in teacher-determined ways and at times of their choosing.
* *One or more Through-Year Common PTELs (Common PTELs),* to supplement the End-of-Year summative assessment in grade five, grade eight, and high school, the intention being to both contribute to measurement and positively influence instruction. These Common PTELs would be relatively brief tasks, likely one to three class periods in length, that involve students in designing and conducting inquiries, observing and collecting evidence, and drawing conclusions about scientific phenomena.
* *End-of-Year summative assessment that includes a section of discrete items plus a Mini PT,* a brief task of perhaps 20 to 30 minutes, potentially a simulation, that draws on the inquiry skills students are developing in the classroom.
* *OTL Indicators,* to both provide context for interpreting scores and guidance for strengthening instruction.

Figure 1 presents a schematic representation of how these four components are arranged into an assessment system (with this proposal’s two focal components highlighted by orange dots). These four components, described in more detail below, will form a cohesive system focused on serving instruction and state summative assessment purposes. Each Through-Year PTEL, whether a Classroom PTEL or Common PTEL, is constructed upon the same CA NGSS foundation. Synergy between the Common PTELs and Classroom PTELs could be engineered through overlapping contexts, modes of inquiry, and/or data sets that may foster better integration of teaching, learning, and assessment. By design, the Through-Year Classroom PTELs also give teachers the flexibility to adapt materials to best support students’ sensemaking of phenomena by providing personalized and student-centered learning opportunities.

**Figure 1**

*A New Conceptual Approach for an Assessment System*



***Note*.** IMPACT focuses on elements 2 and 3 of the conceptual approach: Through-Year Common PTELs and End-of-Year assessment with Mini PT*.*

The components of the summative portion of the assessment system (Through-Year Common PTELs and End-of-Year summative assessment) will be tailored to, and administered solely in, grade five, grade eight, and once in grade ten, eleven, or twelve. However, the curated set of Classroom PTELs would be available to teachers in all grade levels as an aid to instruction and classroom assessment.

In addition to coherence among the components of the assessment system, we would take a developmentally informed approach to the science practices and scientific sensemaking to be demonstrated by students over time, such that tasks become more complex as students’ progress through grade levels (aligned to the progressions of the SEPs and CCCs across K–12). The sequence should support students in acquiring the ability to conceptualize and conduct investigations of greater complexity as they mature and take more intellectual agency for asking and addressing questions using scientific ideas and practices.

Through this CGSA grant, we propose to develop, pilot, field-test, and evaluate a set of Common PTELs for use during the school year in grade five, grade eight, and once in high school, as well as a set of Mini PTs for the same grade levels. Data from these administrations will be merged and matched with that from the operational CAST administration, thereby allowing simulation and evaluation of an End-of-Year assessment that includes the Mini PT. These Common PTEL and Mini PT prototypes will be constructed with connection to the bank of tasks being developed for classroom use (Classroom PTELs), although that bank will be developed under separate auspices, building on the work that many states are jointly undertaking as part of the State Performance Assessment Learning Community. Similarly, the OTL Indicators, which are part of the conceptual framework above, will be undertaken separately, though connected to the kinds of science learning opportunities envisioned by these tasks.

While this scope of work will take on two of the four components shown in Figure 1 (Common PTELs and End-of-Year assessment with Mini PT, as indicated by the dotted lines), we discuss all four components here to illustrate how the system will cohere.

*Through-Year Classroom Performance Tasks.*The first component is a curated set of PTELs intended for classroom use in all grade levels to support instruction and formative assessment at the teacher’s discretion any time during the year. These Classroom PTELs are envisioned to range in design, format, scope, and duration. The PTELs include activities that draw upon various performance expectations integrating selected SEPs, DCIs, and CCCs and, as such, should help students achieve the standards in the CA NGSS. The tasks will also be designed to help students engage with the SEPs—for example, developing and using models, planning and carrying out investigations, and engaging in argument from evidence.

The flexibility of these Classroom PTELs is a key feature, allowing teachers and students to choose, and teachers to adapt, to best suit students’ diverse learning needs. That same flexibility also makes their use for summative assessment infeasible because of challenges with generating comparable scores across the many different variations of the tasks teachers may generate.

As noted, these Classroom PTELs will build on existing resources being developed within and across partner states and coordinated in an accessible task bank, tagged to standards, task types, and grade levels for ease of use.

Among existing resources that might be drawn upon are the following:

* [ClimeTime](https://www.climetime.org/assessments/resources/) (ClimeTime Assessment Project, n.d.) is a resource in Washington State that focuses on climate-science education. It offers a range of assessment tasks and instructional materials related to climate science. Teachers can access the ClimeTime tasks to find assessments, performance tasks, and other resources designed to help students learn about climate-related topics. These materials are tagged to educational standards and can be integrated into classroom instruction to support learning in the context of climate science.
* The [Assessment Resource Bank](https://www.cde.state.co.us/assessment/resourcebank-assessments) (Colorado Department of Education, 2023) provides teachers in Colorado with a wide variety of assessment items, rubrics, and performance tasks. Educators can use these resources to assess student understanding of state standards and curriculum. The resource bank offers a searchable database of assessments and instructional materials to support teachers in their assessment and instructional planning.
* [Stackable, Instructionally-Embedded, Portable Science (SIPS) Assessments](https://sipsassessments.org/) (SIPS Partners, 2021) is a consortium of eight states, led by Nebraska. It aims to develop a collection of resources, including performance tasks and instructional materials. Nebraska, for example, has developed a Science Classroom Formative Task Repository. This repository includes performance tasks tagged to specific standards and instructional units. The tasks in this repository are intended for formative assessment and classroom use, helping teachers gauge student understanding as they progress through a science curriculum.
* The [Performance Assessment Resource Bank (PARB)](https://www.performanceassessmentresourcebank.org/) (PARB, n.d.) is a nationally recognized source for performance assessment tasks. Teachers can access a wide range of tasks, rubrics, and scoring guides across various subjects and grade levels. PARB’s resources are designed to support educators in creating assessments that align with academic standards and promote deeper learning.
* [Contextus.Science](https://contextus.science/) (Contextus.Science, n.d.) is an assessment-resource hub developed and maintained by the Council of Chief State School Officers’ Science Collaborative and the Council of State Science Supervisors. States can access a full range of shared assessment tasks across K–12, along with associated practice briefs, development and use tools, and supports for building educator capacity in high-quality science performance assessment.

Among these tasks are some that will serve as good models for the Through-Year Common PTELs, discussed next, which are a focus of development for this grant.

*Through-Year Common Performance Tasks.* The second component will be a curated set of Common PTELs, which can be used in a through-year fashion within the state testing window to supplement the end-of-year CAST, thereby contributing to the overall summative score. These Common PTELs would represent standardized versions of the Classroom PTELs, intended for administration without modification, unless such modifications were state approved, for example, to accommodate EL students or students with disabilities.

The Common PTELs will closely align with the CA NGSS performance expectations for the testing grade band and will purposefully represent key elements of the summative-assessment blueprint, thereby making possible the combination of results with students’ End-of-Year summative score(s). In grades five and eight, all students within each testing year will complete one Common PTEL per occasion (where the number of occasions could range from one to two or three, yet to be decided by the state). In particular, the Common PTELs will focus on aspects of state standards that are particularly challenging to assess with shorter assessment tasks. This focus will include such aspects as authentic engagement with the Science and Engineering Practices (SEPs), together with conceptual ideas, in service of meaningful sensemaking, including

* asking questions;
* developing and using models;
* planning and carrying out investigations;
* analyzing and interpreting data;
* using mathematics and computational thinking;
* constructing explanations;
* engaging in argument from evidence; and
* obtaining, evaluating, and communicating information.

Measuring the above SEPs via the authentic representation of behaviors in which scientists engage as they make sense of the natural and designed world requires that students have opportunities to engage in sustained and often iterative practice, rather than performing the kinds of isolated tasks that are most easily accomplished in selected-response contexts. In addition, emphasizing a diverse set of science practices that transcends language barriers will allow varied ways to demonstrate scientific understanding to address diverse students’ needs. For example, instead of relying solely on written explanations or verbal presentations, the performance tasks may require students to conduct experiments, analyze data, or create models to demonstrate their understanding. This approach ensures that all students can engage meaningfully in scientific sensemaking regardless of their language background and promotes inclusivity by valuing diverse forms of expressing understanding and reasoning. The Common PTELs, along with the End-of-Year assessment with Mini PTs (described below), will allow CAST to generate more valid scores while providing better support to educators seeking models of high-quality science teaching and learning.

For each Common PTEL administration, design possibilities could include administering the same PTEL to all students in a grade level; spiraling multiple parallel versions across classes or schools to increase reproducibility; or spiraling PTELs that differ in the SEP, DCI, or CCC to maximize blueprint coverage. Decisions will need to be made regarding the domain on which the PTELs will focus within a given year and grade level so that balance of content coverage across domains is taken into account. These decisions will be made during the initial phases of the CGSA work, informed by input from the field and careful consideration of how Common PTELs can enhance what is being measured by CAST.

Tasks will be designed to be developmentally appropriate examples of what “doing science” looks like. For example, for students in grades ten through twelve, the Common PTELs are envisioned to be more sophisticated performances: they will follow CA NGSS progressions that expect high school students to engage in increasingly student-driven, inquiry-based tasks when compared to the earlier grade levels, resembling tasks requiring student agency and decision-making, like those in the Advanced Placement® (AP) Computer Science Principles examination, which is described below.

The Common PTEL(s) will likely be designed to be completed within one to three class periods in grades five and eight, and perhaps a somewhat longer duration at the high school level, where students may design and conduct a more robust investigation, prototype a design solution, or engage in iterative modeling and argumentation. These Common PTELs may include technology-supported tasks, hands-on activities with documentation of student work captured digitally and uploaded, or combinations of those forms. Additionally, the Common PTELs could involve collaborative portions, with the possibility of scoring the contributions of the group, the performance of individuals, or both.

There are several assessments that may be useful as models for developing the Common PTELs. These assessments were created to measure student competency in various domains and include research prototypes as well as operational programs. Among them are the following:

* Cognitively Based Assessment of, for, and as Learning (CBAL) science performance task prototypes: The CBAL(Liu et al., 2013) science performance task prototypes were designed to assess student understanding in through-course fashion. This performance task prototype was developed for formative assessment purposes and designed to be administered during a two- to three-week instructional period.
* New Hampshire’s Performance Assessment of Competency Education (PACE) program common tasks: [New Hampshire’s PACE](https://nhdoe.instructure.com/courses/58/modules#module_87) (New Hampshire Department of Education, 2023) program included both common assessment tasks and tasks designed by each participating LEA’s staff to align with state standards. The PACE tasks were designed to be multistep, curriculum-embedded assignments that would typically take place over more than one class period.
* Washington State’s [ClimeTime](https://www.climetime.org/assessments/resources/) Assessment Project, which was also described as an example resource for the Classroom PTELs, provides teacher guides, task facilitation slides, student task documents, and sensemaking tools for educators and learners. As such, it could serve as a model for the supports and resources required for developing shared standards for a common task across schools and LEAs. The tasks could be completed in one class period, although many are modularized and could be extended over multiple periods, as needed.
* Many APexaminations have one or more performance tasks, some of which are especially notable for the degree to which those tasks reflect activities that constitute the essence of domain performance. Among those examinations are [Computer Science Principles](https://apcentral.collegeboard.org/courses/ap-computer-science-principles/course), [Seminar](https://apstudents.collegeboard.org/courses/ap-seminar), and [Research](https://apstudents.collegeboard.org/courses/ap-research/assessment) (AP Central, n.d.-abc). To varying degrees, these examinations each require students to demonstrate their understanding and skills in activities that are conducted during the year. A similar model might be used for the Common PTELs, particularly for students enrolled in grades ten through twelve, where an extended project, research report, and/or culminating presentation could constitute both the learning experience and its assessment.
* Several international assessment systems in science include components similar to the Common PTEL. In Queensland, Australia, science performance tasks are developed by schools according to the requirements specified in the curriculum and are referred to as “internal assessments.” An internal assessment within a given subject area is composed of such tasks as a data test (approximately 60 minutes plus 10 minutes perusal) or student-led experiments and investigations (approximately 10 hours of class time). The internal assessment contributes equally to students’ summative scores in conjunction with external assessments that are developed and marked according to the Queensland Curriculum and Assessment Authority procedures (Queensland, n.d.). The Singapore–Cambridge General Certificate of Education Ordinary-Level examination includes a component called the School-based Science Practical Assessment (SPA) for the physics, chemistry, and biology subject areas (Singapore Examinations and Assessment Board, 2023). A total of one hour and 30 minutes is allocated to students to complete the SPA, which may consist of one or two question-sets for each science subject area.
* Several formerly administered state science assessments, such as those of Connecticut and Vermont, may also provide examples for Common PTELs.

As with the Classroom PTELs, the Through-Year Common PTELs should be based on the CA NGSS–aligned grade-level performance expectations, provide feedback to both students and teachers, and promote the development of deeper learning in science.

*End-of-Year Summative Assessment with a Mini PT.* The third component of the conceptual approach is an End-of-Year summative assessment. As in the current CAST, the assessment would include a section of discrete items. Unlike CAST, this End-of-Year component would also include a Mini PT designed to align with state standards and with key features of the Common PTELs, such that students’ access to the task is supported by their earlier classroom work. Commonalities might include a focus on aspects of inquiry such as modeling, analyzing data, or constructing explanations from evidence. There might also be shared data sources, contexts, or phenomena.

These tasks would be streamlined and logistically suitable for an End-of-Year accountability assessment. The Mini PT would most likely be administered within a single class period and responded to digitally. It would be an inquiry task that presents a scientific phenomenon, engages students in scientific practices, and supports them in developing their responses.

The Mini PT could be designed to resemble the scenario-based tasks or simulations successfully used on such instruments as the National Assessment of Educational Progress (NAEP) science assessment (NAEP, 2019) and the NAEP technology and engineering literacy assessment (NAEP, 2018). For example, many technology-enhanced tasks used on NAEP involve student interactions with a simulated phenomenon or problem, decision-making about experimental design, original analysis of data from the simulation, and/or making sense of the data relative to the original phenomenon or problem.

Furthermore, the NAEP Science Framework beginning in 2028 includes explicit attention to the range of lived experience, cultural and linguistic backgrounds, and engagement and motivation factors for learners taking the assessment (National Assessment Governing Board, 2024), addressing the student-centered criterion California has called for. New NAEP science assessments will include

* authentic phenomena and problems grounding all assessment tasks, with attention given to contexts that focus on legitimate interests of specific communities;
* provision of sufficient background information via multiple modalities to support student engagement with tasks that are further from their own lived experience; and
* tasks that allow for critical examination, via the science practices and appropriate disciplinary core ideas, claims, data, and explanations for phenomena and problems.

Simulation tasks are also being piloted by various states, such as Massachusetts, where the science assessment system connects on-demand, simulation-based sensemaking tasks to rich classroom performance assessments. The goal is to provide teachers with better information about what students know and can do than might be possible via a selected-response assessment alone. The on-demand tasks ask students to engage in a simulation of a phenomenon, observing and recording outcomes, drawing inferences about the relationships between aspects of the phenomenon and the outcomes of interest. The system design uses “bridging” to connect the performance tasks to the on-demand simulation-based tasks: the performance tasks themselves are designed to mirror features of the on-demand tasks that educators have identified as requiring practice for students.

Similarly, Connecticut’s former Connecticut Academic Performance Test included classroom-based performance tasks in which students were asked to conduct a defined investigation of modest scope, collect and organize evidence from their observations, and draw conclusions about the problem. The investigative work was collaborative; the final write-up was individual. These efforts then became the basis for a performance item on the End-of-Year assessment, which might reprise results of a group’s work and ask for an analysis of the data to draw conclusions, or an evaluation of the appropriateness, replicability, or generalizability of the inquiry design or process. In this way, the performance task was educative and supported success on the End-of-Year assessment while also contributing to the evaluation of student learning.

*Opportunity-to-Learn-in-Science Indicators*. The fourth component, Opportunity-to-learn-in-science indicators (OTL Indicators), will also inform ongoing improvement in science curriculum, instruction, and supports. Although not part of this competitive grant application, the OTL Indicators are a critical part of the conceptual approach, intended to reflect the extent to which students have equal access to the resources needed to achieve the CA NGSS in school. The OTL Indicators are not intended to be used in scoring but only to provide contextual information to facilitate the interpretation of summative scores and to guide decisions concerning such considerations as state investments, curriculum, and professional development.

Just as the NAEP science assessment includes a survey that allows teachers and students to share information about their resources and curricular experiences, the OTL Indicators will allow for equity and quality analyses that illustrate both the range of resources and experiences across the state and the relationships between various resources, experiences, and outcomes. The inclusion of these indicators may foster a better understanding of the factors that affect student science proficiency and offer data actionable at the local and state level for policy and for program planning and evaluation (National Academies of Sciences, Engineering, and Medicine, 2019).

**Theory of Action**

This section describes the theory of action underlying the proposed conceptual approach. The theory is presented as a logic model depicting a holistic assessment design intended to influence instruction and learning, as well as high-level decision-making (see Figure 2). The theory describes how the four assessment components (Through-Year Classroom PTELs, Through-Year Common PTELs, End-of-Year summative assessment with Mini PT, and OTL Indicators) individually and collectively work with the CA NGSS and teacher professional support to improve science teaching and learning. This improvement is theorized to work by

1. enhancing student learning and motivation in science,
2. improving teacher practices,
3. facilitating more effective decision-making by interest holders and policymakers, and
4. increasing community and public awareness regarding equitable learning opportunities and outcomes.

**Figure 2**

*Theory of Action for the Conceptual Approach*



Each of these goals is intentionally served by elements of our assessment system design. For example, students engage in authentic and rigorous science by participating in Through-Year Classroom PTELs, Through-Year Common PTELs, and the End-of-Year assessment with Mini PTs—all action mechanisms in the theory. Engagement with these authentic tasks to make sense of phenomena relevant to students’ interests and life experience is intended to catalyze for students the ability to make sense of scientific phenomena (intermediate outcome), thereby improving student learning and motivation in science (ultimate outcome).

Similarly, timely information allows for the making of more informed inferences and more sensible adjustments to instruction and learning (action mechanisms), leading both teachers and students to use formative assessment more skillfully (intermediate outcome).

The End-of-Year assessment’s section of discrete items, the End-of-Year assessment’s Mini PTs, the Through-Year Common PTELs, and the OTL Indicators serve as tools for informing local or state policymakers and community members (action mechanism). Policymakers can, as appropriate, identify needs and provide necessary support to LEAs, schools, and classes, and community members (intermediate outcomes).

Finally, teachers emphasizing instruction around the CA NGSS SEPs (action mechanism) should result in improved understanding of the CA NGSS, increased responsive pedagogical knowledge, and greater assessment literacy (intermediate outcomes). These intermediate outcomes, in turn, should lead to the ultimate outcome of enhanced teacher practices and, through that, to improved student learning and motivation in science.

Implementing this conceptual approach poses various challenges and barriers that need to be carefully navigated. Challenges include financial constraints, inevitable variability in classroom implementation, variation in school resources, and variation in the level of support provided by counties or LEAs.

Careful consideration of essential supports is also crucial to successful implementation. These supports include appropriate state and local policies; systems for continuous monitoring, evaluation, and refinement; and sufficient infrastructure for data systems and reporting.

**Project Goals, Objectives, Activities, Timeline, and Outcomes**

The CDE and SBE are committed to advancing a more instructionally supportive assessment system. Given the highly ambitious scope of the conceptual approach, this project will focus as noted on two summative assessment components: the Common PTELs and a related End-of-Year Mini PT, as part of the state’s End-of-Year assessment. Other components in the conceptual approach will be pursued through other funding channels. The state will seek additional funding for some components within the IMPACT scope, such as participation fees and consulting fees, ensuring equitable access and participation across all interest holders. The CDE and SBE plan to collaborate closely to ensure coherence and synergy across facets of the new conceptual approach.

*Project Goals*

The goal of this project is to develop, pilot, field-test, and evaluate a comprehensive set of Through-Year Common PTELs for grades five and eight and one grade level in high school, along with a related set of Mini PTs.

*Project Objectives, Activities, and Timeline*

In line with the above goal, the project objectives include the following: (1) develop Common PTELs and Mini PTs, (2) develop scoring strategies, (3) iteratively analyze the performance tasks, and (4) provide professional development services to educators. Each of these objectives is described in more detail in the following sections.

**Development of Performance Tasks.**

The project will develop a small set of Common PTELs and related Mini PTs aligned with the CA NGSS for grade five, grade eight, and high school. These tasks will share a small set of DCIs and associated learning progressions across different grade levels. This strategic design not only ensures consistency and coherence in assessment practices but also facilitates a seamless transition in evaluating science competency as students advance through levels of scientific proficiency. The task development will take place iteratively throughout the project duration, allowing for continuous refinement and enhancement to best meet the evolving needs of students and educators.

The design decisions for the Common PTELs and Mini PTs are critical to the success of the project. Design considerations include factors such as duration in class time, integration of CA NGSS dimensions, the role that teachers should play in administration and scoring, allowable response modalities, and key features for inclusion in scoring rubrics. Planning activities focused on design decisions should result in a blueprint depicting the domain coverage for the set of tasks to be completed and a set of design specifications detailed enough to guide developers in task creation. The blueprint will be especially important in ensuring that a diverse set of Common PTELs and Mini PTs is created and that they afford domain coverage sufficient to allow some degree of fine-grained, group-level scoring than just a single overall metric.

Activities related to design decisions will be conducted by the CDE and SBE through collaborating with the inquiry Hub team at the University of Colorado Boulder, led by William Penuel, professor of learning sciences in the School of Education. His responsibilities will include working in collaboration with the CA NGSS Collaborative and other California-based science education organizations to gather input from the field on the features and design of Common PTELs and the Mini PTs. This collaborative effort will ensure that the resulting performance tasks meet the needs and expectations of educators and align closely with the project goals.

**Development of Scoring Strategies.**

In addition to traditional summative scores, the project recognizes the value of incorporating performance tasks to provide a deeper and more comprehensive measure of scientific proficiency, reveal nuances in students’ abilities to apply scientific concepts in novel situations, and offer insights into specific areas of strength and areas for improvement. The project will explore scoring methodologies to effectively combine scores from the Common PTELs and Mini PT with the End-of-Year discrete sections.

In addition, the project will leverage AI techniques to explore innovative scoring models for constructed responses, with the goal of providing meaningful information and insightful feedback to enhance learning and teaching. AI-powered scoring models can, in theory, analyze complex responses and provide detailed feedback on students’ performance, offering personalized recommendations for improvement. Leveraging ETS’ prior research on automated diagnosis of students’ reasoning patterns associated with the three dimensions of science learning (Liu et al., in press), this project will use data collected from the pilot and field studies to build and validate AI models. These models may be able to automate the scoring of students’ constructed responses and delineate various reasoning patterns so that qualitative characterizations of those patterns can be reported to teachers and students (ETS, 2021). Within this project scope, we will explore extended AI scoring with selected segments of the performance tasks.

**Iterative Analysis of Performance Tasks.**

The project will engage in a rigorous process of iterative studies to ensure the validity and reliability of the Common PTELs and Mini PTs through three study phases:

* 1. *Cognitive Labs and Interviews*. In Year 1, cognitive labs will engage students from relevant grade levels to explore the construct validity of Common PTELs and Mini PTs, in particular whether those tasks measure the intended thinking processes and science knowledge. These labs will help ensure task alignment with the CA NGSS and grade-level expectations. A diverse selection of students will participate. Each student will have the opportunity to respond to a set of Common PTELs and Mini PTs presented in storyboard format. This inclusive approach ensures that the performance tasks undergo thorough scrutiny from a range of perspectives, reflecting the diversity of the student population the project aims to serve. Think-aloud data will be recorded, which will allow us to capture a fuller picture of what students did when completing the tasks. Students’ think-aloud data will be transcribed and coded for both construct-relevant and construct-irrelevant thinking. Analysis of the data will provide insights into whether the tasks effectively prompt students to engage in science practices to demonstrate their understanding and offer the opportunity to refine any issues with performance tasks, such as students being distracted by construct-irrelevant features like unclear directions or poorly organized presentations of task information.

In Year 2, cognitive interviews will be conducted with a selection of students (15 students per grade level) from the pilot participants, using the programmed versions of the tasks. This set of interviews will help confirm that the intended practices and knowledge are being tapped.

* 1. *Pilot Test*. In Year 2, initial implementation of the Common PTELs and Mini PTs in a small group of schools (up to five schools with up to 300 students). The invited schools will be selected to create a sample that is diverse in terms of key demographic variables. The pilot will inform revisions to the tasks, human-scoring rubrics, automated scoring and qualitative characterization (e.g., extent of engagement with scientific practices, reasoning patterns), delivery platform, and administration logistics. Teacher surveys will also be conducted regarding perceptions of how well the performance tasks allow students to demonstrate scientific competency. Example questions may include the following: What science practice is targeted in this task? Is there anything unclear in the task presentation? Does the task allow for multiple ways of demonstrating competency?
  2. *Field testing*. In Year 3, a larger-scale field test (up to 20 schools with about 3,000 students) will be conducted to further refine Common PTELs and the Mini PTs and to provide data for a more robust evaluation of automated scoring and qualitative-characterization methodology. To approximate a sample of students representative of California’s public-school population, we will stratify the state’s schools according to key demographic variables (e.g., socioeconomic status, rural/urban) and randomly select schools for inclusion within each stratum. We will then contact the selected schools and replace those that decline to participate with another randomly sampled school from the same stratum.

**Data Analyses and Evaluation of AI Scoring.**

The major planned analyses will include the following. All listed analyses will be conducted with the field test data. Selected analyses (e.g., evaluating task performance to inform revisions) will be conducted with the pilot test data, given sufficient sample sizes.

* *Evaluating task performance to inform revisions*. Typical analyses such as the item difficulty, discrimination, timing, and differential item functioning can be used. Teacher survey responses from the pilot will also be analyzed to inform revisions of the tasks.
* *Exploring innovative ways to use process data such as clickstream, simulation sequence, and response time.* Such data may provide meaningful information about students’ science inquiry skills to inform learning and teaching.
* *Exploring the reporting of subscores on SEPs at group level*. The Common PTELs are intended to engage students in SEPs. Each Common PTEL may measure one or two practices. When a set of Common PTELs measuring different practices are developed and randomly assigned to students, it provides an opportunity to report subscores for the practices at the LEA or state level.
* *Evaluating the psychometric properties of different aggregation methods for the reporting of a summative score*. Multiple components in this new assessment system will contribute to a final index of students’ standing on the science assessment at the end of the year. The scores from multiple components need to be aggregated in a meaningful way to support the interpretation of the scores. A test dimensionality study will first be conducted to evaluate whether the multiple components, particularly the performance tasks and the discrete section, measure the same construct. This analysis will inform whether all items can be calibrated together under the unidimensional item response theory (IRT) model or different components or combination of components (e.g., the Common PTELs and the Mini PTs) should be calibrated separately. Potential aggregation methods under consideration include a single scale score (under the assumption that the test is unidimensional) or a weighted average of scores from each component (Common PTELs, Mini PTs, and the discrete section), or a combination of components (Common PTELs plus Mini PTs, and the discrete section).
* *Evaluating the psychometric properties of the aggregated scores with different discrete-section lengths*. For this analysis, the Common PTEL and Mini PT data collected in this project will be merged with operational CAST data for the same students. These analyses will help identify whether the discrete section can be shortened and by what amount. Psychometric properties can be simulated by combining students’ scores from the Common PTELs and Mini PT with the scores from various lengths of the operational CAST discrete section.
* *Evaluating the accuracy of automated scoring and automated qualitative characterizations of student performance*. The project team will leverage the data collected during the pilot to explore automated scoring approaches. These AI-supported approaches will not only facilitate the efficient scoring of constructed responses but also provide educators with information to inform instructional next steps. Given the budget constraints, we will select one to two constructed-response items to evaluate such methods. For each item, two human raters will use the rubrics to double-score 20% of the student responses per item and to code each response for qualitative-feedback purposes. Goodman & Kruskal’s lambda (Goodman & Kruskal, 1954) and appropriate versions of kappa (Cohen, 1960; Fleiss, 1971) will be employed to measure the agreement of the automated judgments with human judgments. The human-coded data will be used to both train and evaluate the AI methods. A sample size of 200 responses per task will give enough statistical power to estimate the probability of exact agreement within a reasonable margin of error.
* *Investigating the prediction invariance in AI scores and qualitative characterizations across demographic groups to identify any instances of differential functioning*. We will fit multinomial logistic regression models with the machine-predicted qualitative characterizations as outcomes and the human characterizations and subgroup indicators as predictors. To determine whether there are differential prediction errors by, for example, language proficiency, we will first test whether the interaction between human classifications and subgroup indicators is statistically significant with a likelihood ratio chi-squared test. If the interaction is not significant, we will also test the main effect of demographic group while controlling human classifications to determine whether the human judgments completely mediate the relationship between machine scores and demographic variables.

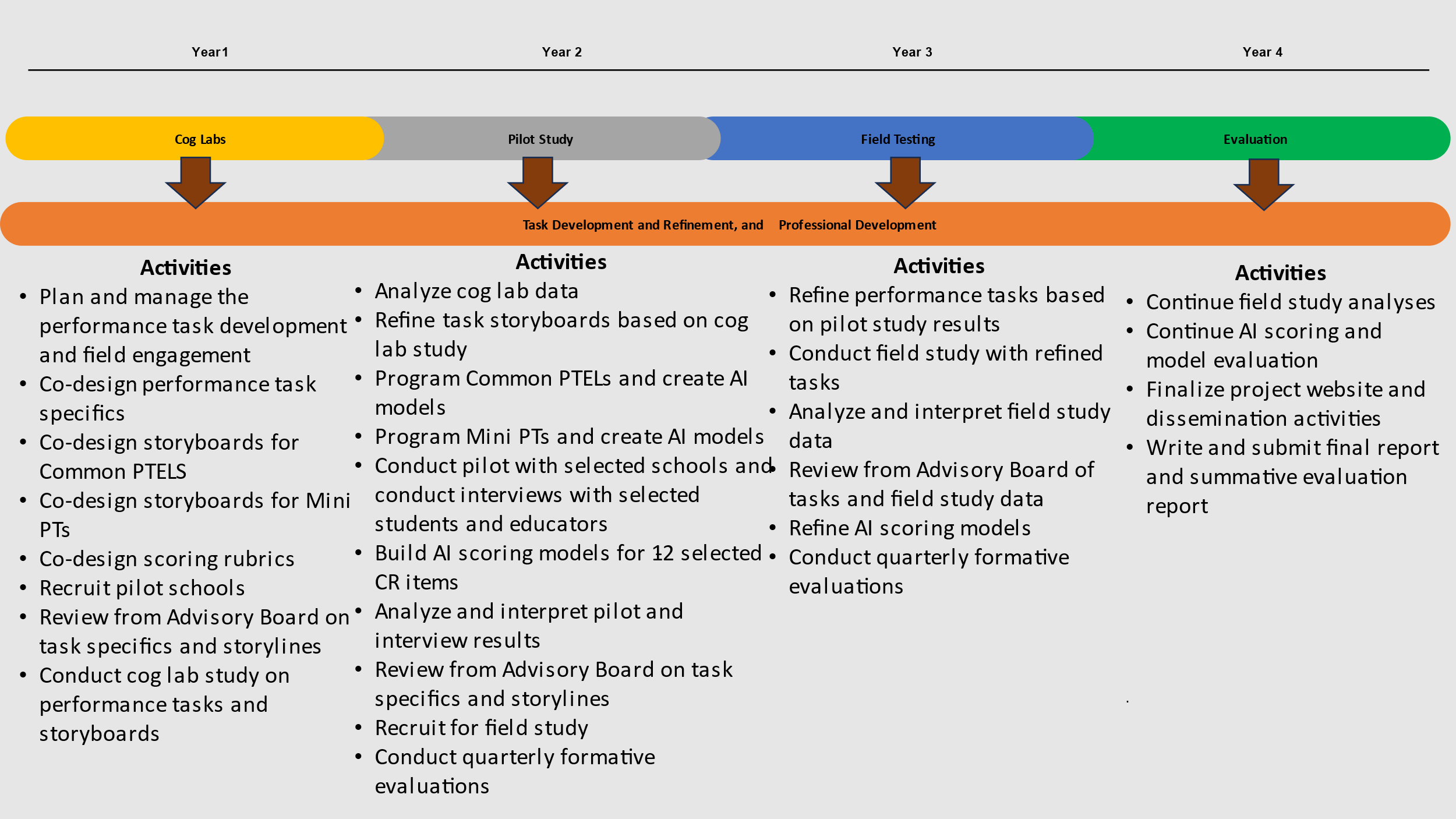
**Professional Development.**

The project will provide professional development and support to educators for piloting, field-testing, scoring, and interpreting results from Common PTELs and Mini PTs. Professional development sessions will be tailored to meet the diverse needs of educators, catering to varying levels of experience and expertise. These sessions will cover a wide range of topics, including assessment literacy, strategies for scoring student responses, and techniques for interpreting assessment results to inform instructional decision-making. In addition, the project will provide access to a variety of resources, including scoring rubrics, online tutorials, and a collaborative learning community where educators can share insights, exchange ideas, and learn from one another’s experiences.

Figure 3 provides a timeline of project activities.

**Figure 3**

*Timeline of Project Activities*



**Project Outcomes.**

The expected outcomes of the project include

1. a finalized set of Common PTELs and Mini PTs ready for potential statewide implementation in grade five, grade eight, and high school;
2. an evaluation of approaches for combining information across the Common PTELs and the End-of-Year assessment for purposes of individual and group score reporting;
3. improved alignment of science assessment practices among project participants with the CA NGSS, emphasizing SEPs and competencies;
4. better and more actionable insights for project teachers into student progress throughout the school year, facilitating targeted instruction and support;
5. increased capacity among participating educators to utilize performance-based assessments effectively, interpret assessment results, and tailor instruction to meet diverse student needs;
6. enhanced validity and reliability of science assessment data, contributing to more informed decision-making; and
7. greater equity of access to high-quality assessment tools and resources for participating students and educators.

## Quality of Project Services

The project design described above is the basis for the implementation of impactful services. These services will encompass a range of activities tailored to ensure equitable access and opportunity, address needs of the intended beneficiaries, provide quality training and professional development services, and generate improvements in practice.

## Strategies for Ensuring Equitable Access and Opportunity

*Further Inclusive Engagement with Interest Holders from Diverse Backgrounds*

Given that California has one of the most diverse student populations, it is imperative to employ strategies that foster inclusive design through collaboration with interest holders from various backgrounds. In addition to California interest holders, other state education officials will be invited to review materials during development. The CDE’s development of the concept paper for this assessment innovation was part of a public process, which initiated considerable interest-holder engagement and enthusiasm. We emphasize that instructional and CA NGSS alignment between the assessment innovations, namely the Common PTELs and the End-of-Year Mini PT, is central to the project and its underlying conceptual approach.

To further engage interest holders in the performance task–development process, the CDE will adopt an inclusive and participatory design approach (Bang & Vossoughi, 2016; Penuel et al., 2020). This approach reflects features of a family of related design and research approaches in being committed to advancing insights about learning and development by attending to what forms of domain-specific knowledge should ideally be generated, while emphasizing how, why, where, by, and for whom such knowledge is generated (Bang & Vossoughi, 2016). Such approaches tend to promote more democratic practices in education systems, equalize power relations between and among researchers and interest holders through continued engagement, recognize that learning is situated in multiple interconnected contexts, and highlight the importance of mutual learning among participants (Cumbo & Selwyn, 2021; Luck, 2018; Penuel et al., 2020). Within the scope of the proposed work, this approach will entail conducting focus groups and interest-holder engagement activities to (1) deepen our understanding of current practices and desired shifts in assessment methodologies, (2) explore the diverse learning experiences of youth within and outside the classroom to help inform task design and balance lived experiences with appropriate expectations for transfer and generalizability, (3) understand what scores and information from the Common PTELs and the End-of-Year Mini PT might be most useful for instruction and student learning, (4) identify possible collaborators for co-design efforts in subsequent stages of this work, and (5) more generally inform the priorities for CAST Common PTELs and Mini PTs.

With respect to informing performance-task priorities, there are multiple ways of thinking about those priorities. One way might concern the SEPs, CCCs, and DCIs that should be highlighted (i.e., to the extent that the CA NGSS does not stipulate a weighting, should some practices, crosscutting concepts, or core ideas be more frequently represented in performance tasks than others?). Interest holders (e.g., business leaders, university professors, science educators, parents) may have valuable perspectives to share on this question. An alternative way of thinking about performance-task priorities concerns how to make those tasks maximally accessible to students from different cultural backgrounds. Community representatives can offer contexts within which to embed PTs so that they connect better with the lived experiences of students from particular demographic groups. Representatives will be able to advise on the benefits and limitations of building student choice or agency into such tasks, that affordance being characteristic of performance tasks in some operational assessment programs, as noted below.

Critical questions will be explored during these interest-holder engagement activities to ensure that the design of Common PTELs and Mini PTs facilitates more equitable access:

* To what extent do educators currently integrate real-world experiences and phenomena into their teaching and assessment practices?
* What challenges do educators face in implementing performance tasks in science classrooms?
* What resources and support do educators need to effectively use information collected from performance tasks in teaching?
* How can assessments better reflect the diversity of students and backgrounds in the local community?
* What types of performance tasks and scenarios would promote deeper engagement with science concepts and practices?
* What are the key considerations for ensuring fairness and equity in science performance tasks, particularly for students from traditionally underrepresented backgrounds?

**Strategies for Performance-Task Development: Design Principles and Decisions**

*The IMPACT project* will strive to create authentic, sensemaking performance tasks that are responsive to and reflective of the great diversity of California’s school population. Bennett (2023) suggests five design principles intended for such purposes, four of which directly concern development.

1. Present reasonably realistic problem situations that connect to learner experience, culture, and identity because students are more likely to be able to show what they know in familiar as opposed to foreign contexts.
2. Allow for multiple forms of representation and expression in problem stimuli and in responses, as some forms may be more common to the ways of knowing and community practices of particular groups.
3. Promote instruction for deeper learning through assessment design, which is especially important for less-experienced teachers and schools with fewer curricular resources available to support such learning. In addition to calling for performance tasks, deeper learning implies access to consultative resources similar to ones used in real-world problem-solving and encourages habits of mind like those proficient performers employ (e.g., evaluating their work against quality standards). Making these characteristics part of the assessment design may encourage teaching and learning practice that is more equitable, more like what students in well-resourced schools are likely to receive routinely.
4. Adapt—or personalize—the assessment to student characteristics. Such personalization already occurs with respect to competency level through computerized adaptive testing, to special needs via accommodations and modifications, and to interests and prior knowledge by sometimes affording problem choice, even locally negotiated choice. The latter type of choice pertains when students design the problems to be addressed (within constraints and with teacher guidance), as in AP Research and AP Computer Science Principles (AP Central, n.d.-ab).

In addition, the strong emphasis on a broad spectrum of science practices within the performance tasks serves to promote equitable access for all students, including students from diverse linguistic backgrounds, enabling them to demonstrate their knowledge and abilities in varied and inclusive manners. For example, tasks that involve designing and carrying out experiments, data analysis and interpretation, or constructing and using models, can benefit EL students who may struggle with written or verbal expression but excel in visual and hands-on learning. This design approach ensures that all students, regardless of language proficiency, can actively engage in scientific inquiry and effectively demonstrate their scientific competency.

Overall, these design principles promote equitable access by recognizing and addressing the unique needs and experiences of all students, ensuring that assessments are fair, inclusive, and relevant, allowing every student to demonstrate what they know and can do.

*Strategic Review to Promote Equitable Access*

During task development, to ensure quality and fairness, four types of review will be undertaken: (1) content, (2) editorial, (3) bias and sensitivity, and (4) accessibility. Content review will be conducted to ensure that construct measurement is valid at the storyboard stage. The content review will be done by a demographically diverse group of science educators and assessment specialists. This review will ensure the following:

* Each phenomenon represents an authentic situation or a real-world setting as a context at the outset of a task.
* Students are presented with clear and meaningful purposes for the activity or problem to be solved.
* Scenarios and contexts are used to boost engagement and connection with students’ lived experiences.
* Scenarios, contexts, and problems are accessible to diverse student populations and/or provide sufficient background information.
* The scenario or context is grade appropriate, important, and relevant, including stimuli, sources, images, data, and language.
* Task interactions help assess both NGSS core ideas and practices.
* Items and interactions help maximize collection of evidence about various aspects of student science competency.

Editorial review will apply accessible language standards in items, focusing on clarity, simplicity, and correctness. The bias and sensitivity review will identify and eliminate or revise questions containing content (e.g., pictures, graphics, symbols, phrases, or other references) that could be construed as offensive to or biased against members of specific groups. Trained team members will review each storyboard to ensure that it does *not*

* use language that stereotypes with regard to gender, race, culture, ethnicity, class, or geographic region;
* contain language not commonly used in most areas or that has different connotations in different regions;
* discriminate in any way against individuals with disabilities;
* have references to religion;
* reflect gender or ethnic bias;
* make assumptions that all students are from the same socioeconomic group;
* show anyone in a stereotypical manner;
* use any offensive or demeaning words, symbols, graphics, pictures, phrases, or artwork; and
* contain any emotionally charged subject matter.

The goal of accessibility review is to align performance-task designs with the international standards for digital accessibility, specifically the Web Content Accessibility Guidelines (WCAG), version 2.0. This review involves assessing the tasks for compliance with the guidelines, which cover such accessibility criteria as text alternatives for nontext content, keyboard accessibility, and compatibility with assistive technologies. Such alignment ensures that digital tasks are accessible to all individuals, including those with disabilities, thus promoting inclusivity and equitable access.

*Dissemination Strategies*

Findings and activities of this project will be disseminated through diverse channels and to various audiences, including science educators, school administrators, professional developers, researchers, and the public. The dissemination will occur through publications, reports conference presentations, and a dedicated project website. The project website will serve as a central hub for teacher professional support, resources, and information about the project. These resources will be freely accessible to all teachers during and after the project duration, facilitating their ongoing use and implementation in diverse settings.

**Services Provided Are Appropriate to the Needs of the Intended Beneficiaries**

As shown in the theory of action (Figure 2), there are four overarching goals of the new conceptual approach to the CAST program. These goals are (1) enhancing student learning and motivation in science, (2) improving teacher practices, (3) facilitating more effective decision-making by interest holders and policymakers, and (4) increasing community and public awareness regarding equitable learning opportunities and outcomes.

As noted, the scope of IMPACT will focus on the Common PTEL and Mini PT components of the conceptual approach. Within that scope, the project will provide appropriate services to three groups of beneficiaries: students, educators, and interest holders.

For students:

* Common PTELs and Mini PTs will offer engaging, authentic experiences aligned with the CA NGSS, thereby fostering deeper understanding and motivation in science learning. The focus on the engagement of science practices will help connect science learning to tangible, real-world applications, thus developing a stronger appreciation for the relevance and significance of science in students’ everyday lives.
* By focusing on real-world phenomena and inquiry-based learning, these assessments should promote active participation and meaningful learning experiences for students of diverse backgrounds and needs. Immersing students in real-world scenarios and scientific phenomena helps ignite students’ curiosity and encourage active exploration of the natural world.
* The innovative scoring approaches we plan to explore for the Common PTELs and Mini PTs have the potential to provide useful information to help students identify strengths, areas for improvement, and pathways for further progress. The Common PTELs should provide more timely feedback and actionable insights to empower students to regulate their learning.
* The asset-based design and scoring approach will encourage students to recognize the value of their own ideas and thinking approaches as a stepping stone toward more sophisticated levels of scientific understanding (see Fine & Furtak, 2024). By using students’ prior knowledge and their unique perspectives as assets, these tasks will promote a positive experience where all students feel valued and capable of achieving success. This asset-based approach should enhance students’ confidence in their ability to tackle complex scientific challenges (see Fine & Furtak, 2020; Shepard et al., 2020).

For educators:

* Common PTELs and Mini PTs will provide valuable tools and resources to improve teaching practices and curriculum development. These assessments will offer a model for educators to design engaging and meaningful classroom assessments and instructional activities that can promote deeper understanding of scientific concepts and better application of science practices.
* By aligning assessment with the CA NGSS, educators should be able to make more informed instructional decisions and tailor learning experiences to meet the diverse needs of students. Such alignment ensures that the assessments accurately reflect intended learning outcomes and provide educators with valuable insights into their students’ strengths and areas for growth.
* Access to these performance tasks and relevant professional development opportunities supports educators in implementing effective instructional strategies grounded in student-centered learning and formative assessment (Heritage & Wylie, 2020). Access to, and experience with, these performance tasks, scoring rubrics, and training sessions will equip educators with the necessary knowledge and skill to effectively use assessment data to inform instructional practice.
* The potential of AI-scoring should help provide timely and actionable information for educators to adjust instructional planning.

For interest holders:

* Including relevant interest holders in task development will increase transparency and promote trust and confidence. Involving representatives from diverse backgrounds in task development, including educators, administrators, parents, community leaders, and industry professionals, will foster collaboration and ensure the performance tasks are relevant, meaningful, and responsive to the needs of all interest holders.
* Including interest holders in the review of data and results will enable them to make evidence-based and data-driven decisions, guide resource allocation, and drive policy changes. This involvement empowers interest holders to allocate resources strategically, prioritize initiatives, and advocate for policies that support equitable access to high-quality science education.

*Quality of Professional Development*

Professional development (PD) services play a crucial role in offering educators the opportunity to develop a deeper understanding of performance tasks and their role in assessing student learning, as well as informing learning and teaching. The CDE and SBE are committed to supporting educators in effectively implementing teaching and assessment of the CA NGSS. For example, the CDE has organized a series of CA NGSS Rollout workshops and institutes across the state to provide educators with comprehensive training on CA NGSS implementation, including aligned assessment practices. In another initiative, CDE supported LEAs in piloting CA NGSS–aligned instructional materials, assessments, and professional development resources. As part of this latter initiative, educators received ongoing training and support to develop and implement CA NGSS–aligned performance tasks that measure student proficiency in three-dimensional science learning. Furthermore, various CA NGSS collaboratives and networks have been established in the state to facilitate collaboration, knowledge sharing, and professional learning among science educators. These collaboratives bring together educators from different LEAs and schools to discuss CA NGSS assessment strategies, share resources, and engage in peer-to-peer support.

Building off the state’s existing professional development resources and platforms, this project aims to further enhance professional development opportunities focusing on using performance tasks to support CA NGSS assessment, learning, and teaching. The project will identify successful PD practices and explore examples of engaging, interactive, and research-based methods that have resulted in improved teaching and student outcomes. Based on a review by Darling-Hammond et al. (2017), to the extent feasible within the constraints of the project, we will apply the following elements of effective teacher professional development programs to guide our design of virtual PD sessions:

* **Content focus:** The PD will intentionally focus on the integrative nature of science learning (i.e., integrating DCIs with SEPs) situated in teachers’ classrooms practices and pedagogies. Specifically, the PD will support teacher design and use of NGSS-aligned performance tasks within their classroom contexts.
* **Active learning:** The PD will prioritize interactive and participatory approaches, engaging educators in hands-on activities, analysis of authentic student artifacts, and trying out new pedagogical strategies to provide highly contextualized and personalized learning experiences.
* **Collaboration:** The PD will foster collaboration by creating an online professional community where educators can share ideas, resources, and best practices related to NGSS-aligned performance tasks. Through this collaborative platform, educators will have the opportunity to engage in ongoing dialogue, exchange insights, and learn from each other to enhance their collective expertise in using performance tasks effectively.
* **Models of effective practice:** The project’s PD activities will provide educators with exemplars and models of effective practice in designing and implementing performance tasks. Educators will have access to sample tasks, detailed assessment specifications outlining alignment with the CA NGSS, scoring rubrics, sample student responses, and classroom use cases. These resources will help educators gain practical guidance and inspiration for designing, adapting, and implementing performance tasks that effectively support student learning and engagement in science practices.
* **Coaching and expert support:** The PD will include opportunities for personalized coaching and expert support tailored to educators’ individual needs and contexts. Experienced facilitators will provide guidance, feedback, and mentoring to educators as they navigate the process of designing, implementing, and assessing performance tasks. This individualized support will help educators refine their practice to support student science learning.
* **Feedback and reflection:** The PD will provide sufficient time for educators to engage in reflection, receive feedback on their practice, and make iterative improvements. Through facilitated reflection activities, educators will have opportunities to critically examine their practice and collaboratively improve with peers.
* **Sustained duration:** The PD will provide educators with ongoing support over an extended period. The PD will be structured to allow for continuous learning, practice, implementation, and reflection upon new strategies that facilitate changes in practice. Such sustained support will enable educators to internalize the learning and integrate it over time.

**Leading to Improvements in Practice**

The project will generate deeper understanding of the CA NGSS, as well as enhanced science assessment practices and greater assessment literacy, through its inclusive design approach and broad dissemination plans. These intended changes in practice will potentially be further amplified by synergies with initiatives targeting other components of the new conceptual approach (e.g., Classroom PTELs).

*Deeper Understanding of the Science Standards*

By adopting an inclusive design approach, the project ensures that diverse perspectives and voices of interest holders are incorporated into the development of the Common PTELs and Mini PTs. This approach will foster a deeper understanding of the CA NGSS with an emphasis on engagement of science practices. Through collaborative efforts, educators will develop a deeper understanding of the underlying principles and expectations of the CA NGSS.

Other initiatives focused on embedding Classroom PTELs into curriculum development and instructional strategies will provide essential context and support for educators to implement the Common PTELs and Mini PTs and, more importantly, to align assessment practices with their broader science instructional goals aligned to the CA NGSS.

*Enhanced Science Assessment Practices and Assessment Literacy*

By involving educators and school administrators in the co-design and implementation phases, interest holders will have access to high-quality PTELs that accurately measure student knowledge and practices in science. These performance tasks will potentially provide educators with valuable insights into student learning, progress, and mastery of science competencies, enabling more data-informed instructional decisions.

Moreover, the project’s broad dissemination plans will spread knowledge and best practices related to high-quality science performance assessments. Through such channels as reports, webinars, and PD sessions, the project will share its findings and materials with educators, school administrators, policymakers, and the broader community, helping various interest holders enhance their science assessment practices.

The investment in the PD element of the project will contribute to increasing assessment literacy among educators. By providing tailored training sessions, resources, and ongoing support, the project will equip educators with the knowledge, skills, and confidence to administer, score, and interpret results from Common PTELs and Mini PTs effectively.

## Adequacy of Resources

**Reasonable Costs**

The CAST conceptual approach is very ambitious. Considering the reasonable costs that can be covered by the CGSA funding with the state seeking additional supplementary funding, the proposed project will focus on the two summative components of that approach: the Common PTELs and the End-of-Year Mini PT. This targeted approach allows for the allocation of resources where they can have the most immediate and impactful outcomes, while also ensuring the feasibility of the project within the available budget. The proposed IMPACT project will leverage existing resources, expertise, and capabilities from the CDE and ETS to reduce the cost and time needed for the proposed scope. The budget of the IMPACT project reflects a sufficient but reasonable allocation of funds in relation to its purpose, activities, design, and potential significance, with the state seeking additional funding to cover some project components such as participation and consulting fees.

The CDE has been at the forefront of initiatives aimed at implementing the NGSS and enhancing science education across the state. In collaboration with educators, assessment experts, and interest holders, the CDE has offered PD programs and created assessment resources to support effective implementation of the CA NGSS. As a partner, ETS brings its expertise in assessment design, psychometrics, and innovative testing approaches to the project. With a long track record of developing reliable and valid assessments, ETS will play a crucial role in the development, piloting, field-testing, and evaluation of the performance tasks to ensure the tasks provide meaningful and actionable information about students’ science understanding and progress. With its longstanding reputation as a leader in assessment development and research, ETS has demonstrated expertise in developing innovative assessments to address evolving needs of education systems, piloting new assessment models, and creating cutting-edge technologies to improve engagement and learning outcomes.

By combining the resources and capabilities of the CDE and ETS, the IMPACT project aims to streamline the development process, reduce costs, and maximize efficiency. This collaboration will enhance the feasibility of the project within the available CGSA budget, and also ensures the quality and rigor of the assessments.

Finally, by making publicly available its project materials, including selected performance assessment tools and scoring rubrics, professional development modules, and supporting resources, IMPACT ensures that the knowledge and insights generated from the CGSA project can be shared widely with educators, policymakers, and interest holders across the country. This approach will amplify the impact of the project and promote potential collaboration and knowledge-sharing among states and educational institutions.

**Number of Proposed Pilot Sites and Participants**

The prototype PTELs and Mini PTs developed as part of this effort will undergo pilot testing to vet feasibility and make necessary adjustments to the design prior to their revision and field testing. The CDE will work closely with ETS to identify three to five schools (up to 300 students) for the pilot test. Pilot schools will be selected to represent as diverse a set of demographic characteristics as possible among those LEAs interested in participating in the project. This sample size will be sufficient to provide the data needed for task revision prior to the much larger field test.

At the time of proposal writing, several sites have expressed interest in participating, including a school in central California, as well as a county office of education in northern California. Furthermore, the California Association for Science Educators, a nonprofit representing science educators, has been a partner during CAST development and may help in identifying or recruiting sites for participation.

In addition to participation incentives, if desired, ETS will provide informal results from the pilot to give educators another source of data for obtaining insight into what their students know and can do.

Field testing will occur following the prototype and pilot revisions. A sample of up to 20 schools and 3,000 students will be recruited. Demographic characteristics will be an important factor in selecting participants. As noted, a stratified sampling approach will be used to better ensure that the participants are representative of California’s public-school population in terms of key demographic variables (e.g., socioeconomic status, rural/urban setting). Table 1 gives the distribution of the state’s students on one such variable, race/ethnicity.

Similar to the pilot test, schools that participate will receive a participation stipend and will be provided some level of information about student performance.

**Table 1**.

*Race/Ethnicity of Students Enrolled in the California K–12 Public Educational System During the 2022–23 School Year.*

| Ethnicity | Percentage of Population |
| --- | --- |
| African American | 4.67% |
| American Indian or Alaska Native | 0.45% |
| Asian | 9.52% |
| Filipino | 2.18% |
| Hispanic or Latino | 56.13% |
| Pacific Islander | 0.41% |
| White | 20.09% |
| Two or More Races | 4.31% |
| Not Reported | 2.25% |
| Total | 100% |

## Management Plan

**Management Structure and Partner Responsibilities**

We have designed a robust management plan to carry out this work. The CDE will organize and carry out activities under the direction of Principal Investigator (PI) and Project Director Dr. Mao Vang (CDE), and Co-PI Vigdis Asmundson (CDE). Dr. Vang and Ms. Asmundson, with ETS Subcontract-PI Dr. Lei Liu (ETS), ETS Senior Advisor Dr. Randy Bennett (ETS), and ETS California Program Director Mary Anne Arcilla (ETS) and CAST Program Director Kelly Bolton (ETS), will form the Project Leadership Team. This team will coordinate multiple project activities to facilitate communication and synergy among the partners. While the Project Leadership Team will have primary oversight responsibility, its efforts will be supported by a larger management team. Having a larger team with clearly defined roles allows for an effective level of project knowledge to accommodate the staff turnover that inevitably occurs during multiyear periods.

Planning will occur iteratively over the project duration, allowing for alterations based on the data being collected during each phase. Meetings every other month will be held between the CDE Project Team and the ETS team, along with other participants specified in this project narrative, including members of the Advisory Committee as listed below. Table 2 presents the proposed management plan.

**Organizational Responsibilities**

The **CDE** will serve as the grant manager and fiscal agent. Leveraging the expertise of its Assessment Development and Administration Division, the CDE will oversee various aspect of the IMPACT project, including quality control of performance-task development, participant recruitment, school selection, and overall project management. The CDE has a dedicated team responsible for grant management, including experienced professionals in grant administration, budgeting, compliance, and program implementation. Over the years, the CDE has secured and managed numerous federal grants from various agencies, including the US Department of Education. The CDE will work with the SBE to provide oversight throughout the project. The CDE will collaborate closely with ETS as co-leads to ensure successful execution.

As the governing body responsible for setting educational policy in California, the **SBE** brings valuable expertise in education policy, program evaluation, and strategic planning, which will contribute to the successful management of this work.

The **SBE** will play a pivotal role in oversight and defining high-level project requirements. It will provide guidance and oversight of the project management team, ensuring alignment with state educational priorities and policies.

With the CDE, **ETS** will co-lead the work as the primary partner. ETS research has a wealth of experience in securing and managing federal grants, including grants funded by the US Department of Education and the National Science Foundation, particularly in the field of educational assessment. Leveraging its prior success with federal grants, ETS is well-equipped to co-lead the IMPACT project, providing invaluable expertise in assessment development, research, and program management. ETS will play a key role in assessment design, research methodology, and project evaluation, collaborating closely with the CDE and SBE to drive innovation and promote equitable access to high-quality science education resources.

The CDE will also partner with the **Institute of Cognitive Science and School of Education at University of Colorado** which will have the key role in task design and the cognitive labs. This will also includemaking design decisions, developing performance tasks for the IMPACT project, and providing direction on cognitive labs. With its extensive experience and expertise in the design, implementation, and scoring of performance assessments aligned with the NGSS, the inquiryHub team, under the leadership of Dr. William Penuel, will play a crucial role in task development, dissemination, and capacity building in California. The team will actively facilitate the CDE in tapping into cutting-edge research and best practices in performance assessment design and development. Furthermore, the team’s involvement in national efforts since the publication of the NGSS positions them perfectly to lead initiatives aimed at developing performance tasks for evaluation purpose and supporting teachers in integrating three-dimensional performance tasks into their classrooms.

**California Alliances and Infrastructure.** The CDE will partner with the NGSS Collaborative to leverage the NGSS Collaborative’s expertise and advice in conducting outreach to the field and conducting professional development training. California benefits from a robust network of partnerships and alliances dedicated to supporting innovative science assessment initiatives like IMPACT. These entities include statewide programs such as the California Science Project, the California Association of Science Educators, California NGSS Rollout, California NGSS Assessment Development Consortium, and California NGSS Collaborative. These alliances may help support IMPACT efforts across various stages of the project. In the performance task–development stage, these alliances can provide access to expertise, resources, and best practices in science education. Educators and administrators involved in these alliances can contribute valuable insight and feedback during the development processes. During piloting and field-testing, these alliances may help facilitate recruitment of schools and participants, and provide logistical support and assistance in coordinating pilot and field testing activities. In the PD stage, these alliances can serve as key partners in disseminating information about training opportunities and help provide educators with ongoing support and resources.

**Table 2**

*Objectives and Activities by Year.*

| Objective and Activity | Year 1 | Year 2 | Year 3 | Year 4 |
| --- | --- | --- | --- | --- |
| Kickoff meeting and project team formation | X |  |  |  |
| *Objective 1: Development of Performance Tasks* | | | | |
| Plan for development goals, design principles, and overall scope | X |  |  |  |
| Co-design and refine performance task specifics | X |  |  |  |
| Co-design and refine storyboards for Common and Mini PTELS | X | X |  |  |
| Internal review of processes to promote equitable access | X | X | X |  |
| Program Common and Mini PTELs |  | X |  |  |
| Refine programmed performance tasks |  | X |  |  |
| *Objective 2: Development of scoring strategies* | | | | |
| Co-design and refine scoring rubrics | X | X | X |  |
| Build and refine AI scoring models for one to two selected CR items |  | X | X | X |
| *Objective 3: Iterative validation of performance tasks* | | | | |
| Recruit cognitive lab participants | X |  |  |  |
| Conduct cognitive lab study on performance tasks and storyboards | X |  |  |  |
| Analyze cognitive lab study data | X | X |  |  |
| Recruit pilot schools | X |  |  |  |
| Conduct pilot on selected schools and conduct interviews with selected students and educators |  | X |  |  |
| Analyze and interpret pilot and interview results |  | X |  |  |
| Recruit for field studies |  | X |  |  |
| Conduct field study |  |  | X |  |
| Analyze and interpret field study data |  |  | X | X |
| *Objective 4: Professional Development* | X | X | X | X |
| *Evaluation activities* | | | | |
| Summative evaluation |  | X | X | X |
| Formative evaluation | X | X | X | X |
| Advisory Committee review | X | X | X | X |
| Dissemination activities | X | X | X | X |

**Management Leadership Team**

*California Department of Education*

Dr. **Mao Vang** will serve as the project director to monitor the development of tasks and provide overall management of grant activities. Dr. Vang is the Director of the Assessment Development and Administration Division (ADAD) of the CDE. Over the past 30 years, she has held various roles that contributed to her knowledge and shaped her leadership. She earned a doctor of education degree from the University of California, Davis. Her dissertation was titled, *Exploring Students’ Experiences with the California High School Exit Examination February 2006 to May 2007*. She has significant experience maintaining administrative and fiscal responsibility over state and federal funds, including proven success managing federal Title I funding. She leads administrative activities of the state for the California Assessment of Student Performance and Progress (CAASPP) and English Language Proficiency Assessments for California (ELPAC), Physical Fitness Test, California Proficiency Program, and High School Equivalency, as well as national and international assessments serving around 5.8 million students, 300,000 teachers, 2,000 LEAs, and 10,000 schools. She directs the annual implementation of more than 20 million tests administered. She oversees multiyear assessment contracts totaling more than $100 million annually. She participated in the transition from paper–pencil tests to computer-based assessments. Under her leadership, the English, mathematics, and science assessments under CAASPP, the ELPAC, their alternate assessment counterparts, along with interim assessments and formative assessment resources were brought onto one technology platform and one reporting system.

**Vigdis Asmundson** will serve as the project coordinator overseeing the design, development, and piloting of tasks. Ms. Asmundson is the Education Research and Evaluation Administrator for the Science Office in ADAD. She holds a Masters in Educational Administration and is a PhD candidate in School Organization and Educational Policy. She has an extensive background as a science teacher in schools with diverse populations.

A full-time **project manager** with expertise in performance assessment will be hired and dedicated to the IMPACT project to facilitate collaboration among various partners to ensure the smooth execution of project activities, manage engagement with interest holders, and oversee the administration of pilot and field tests.

*ETS*

Dr. **Lei Liu** will act as the IMPACT ETS PI to oversee the development of performance-task scoring, exploration of AI-supported scoring models, and design of iterative validation studies. Dr. Liu is a research director leading the K–12 research team in the ETS Research Institute. She is also an adjunct professor at the University of Pennsylvania. She earned a PhD in educational psychology with a focus on learning sciences and educational technology from Rutgers University. Her research interests lie at the intersection of science learning and assessment, learning sciences, and educational technology. She has led multiple federal grants to develop transformative innovations for science, technology, engineering, and mathematics learning, including topics on learning progressions, AI-supported assessment tools, and virtual labs. She has produced more than 70 peer-reviewed publications. She is a member of the editorial board of *Instructional Science* and has served as a reviewer for multiple international conferences, journals, and National Science Foundation merit reviews. In addition to her lead role in research, Dr. Liu has also been a key contributor to support various operational programs at ETS, including the California state assessment programs and NAEP science and mathematics programs.

Dr. **Randy Bennett** will act as senior consultant, advising on design and scoring principles, research questions, and analysis issues. He is Norman O. Frederiksen Chair in Assessment Innovation with the ETS Research Institute. He directed the 2003 NAEP Technology-Based Assessment Project, which laid the groundwork for NAEP’s transition to digital delivery, was the first digital administration of performance tasks to nationally representative samples of US students, and the first such collection that incorporated process data into analysis and scoring. His work focuses on devising assessment approaches that better account for the social, cultural, and other relevant characteristics of students and the contexts from which they come.

*Advisory Committee*

The advisory committee is comprised of top national experts and educators with a wealth of experience and expertise in science education, performance assessment, and policy. The committee will provide invaluable guidance to support, advise, and evaluate the progress of the IMPACT project to ensure successful development and execution of the project plan. Specifically, the committee includes **James Pellegrino** (a Professor Emeritus at the University of Illinois at Chicago [UIC] and a founding Co-director of UIC’s interdisciplinary Learning Sciences Research Institute); **William Penuel** (a Distinguished Professor of Learning Sciences and Human Development department and Director of National Center for research in Policy and Practice at the University of Colorado Boulder); **Richard Shavelson** (Professor Emeritus of Graduate School of Education at Stanford University); **Steve Schneider** (the Senior Director STEM Research and Entrepreneurship at WestEd); **Jill Wertheim** (Director of SCALE Science at WestEd); **Aneesha Badrinarayan** (Director of State Performance Assessment Initiatives at the Learning Policy Institute); and **Bryan Brown** (the Kamalachari Professor of Science Education, and a professor of teacher education).

## Project Evaluation

Formative and summative evaluations will be conducted to ensure achievement of the intermediate and ultimate outcomes shown in the theory of action (Figure 2).

**Formative Evaluation**

Formative evaluation aims to assess the ongoing progress of the IMPACT project to ensure that it is meeting its goals and objectives effectively and to identify any necessary adjustments for improvement. Formative evaluation will be conducted after each major activity (e.g., professional development). Example methods include regular feedback sessions (e.g., via focus groups) with project team members, educators, and interest holders to gather input on the project’s challenges and successes. These sessions will allow for rapid improvements in the next iteration of each activity. Table 3 shows the focal activity, questions, and methods to be employed.

**Table 3.**

*Formative Evaluation Focal Activities, Questions, and Methods.*

| Focal Activity | Questions | Method |
| --- | --- | --- |
| Cognitive Labs | Were the resulting data useful for task and scoring revision? If not, why not?  What changes might facilitate obtaining more useful data? | Interviews with task designers, interface designers, and computer programmers;  Student questionnaires on the cognitive lab experience |
| Pilot and/or Field Test | Were the resulting data useful for task and scoring revision?  If not, why not?  What changes might facilitate obtaining more useful data? | Interviews with task designers, interface designers, and computer programmers  Teacher surveys on the extent tasks allowed demonstration of scientific competency |
| Professional Development (conducted during pilot and field test) | Did teachers feel that PD provided information needed to administer PTELs and Mini PTs effectively?  Did teachers feel that their knowledge and understanding of the CA NGSS improved, especially with respect to scientific practices?  Did teachers feel that their level of assessment literacy increased? | Teacher interviews and surveys on PD sessions and resources; knowledge questions on the CA NGSS |

We will also draw on our project advisory committee, consisting of researchers, science teachers, and other experts to review, advise, and evaluate project progress and quality. Members will participate in formative evaluation at the project level throughout the life of the grant. Example questions include the following:

* How might the design of the PTELs and Mini PTs be improved?
* How might the plans for the cognitive labs, pilot test, and field test be strengthened?
* How might the plans for and delivery of teacher PD be improved?

Findings and recommendations from the formative evaluations will be reported in quarterly intervals throughout the project lifespan. Findings will be shared with the CDE, SBE, and ETS staff members to provide feedback for continuous improvement.

**Summative Evaluation**

The summative evaluation aims to assess the overall impact and effectiveness of the IMPACT project. The evaluation will consist of a listing of project goals and objectives, a judgment for each goal and objective as to its level of completion, and a description of the evidence that justifies each judgment. Table 4 gives the goals, objectives, and measures. The summative evaluation results will be included in the final report.

**Table 4**

*Summative Evaluation.*

**Goal:** Develop, pilot, and evaluate a comprehensive set of Through-Year Common PTELs for grades five and eight and one grade level in high school, along with a related set of Mini PTs.

| Objective | Measures |
| --- | --- |
| Develop Common PTELs and Mini PTs | A number of completed PTELs and Mini PTs developed, with a majority considered of acceptable psychometric quality as judged by the Advisory Board based on field test data.  PTELs and Mini PTs are judged by Advisory Board and by a sample of surveyed participating teachers to be aligned with socioculturally responsive assessment practices and able to elicit evidence of student thinking in ways that value multilingual and multicultural assets.  PTELs and Mini PTs are judged by Advisory Board and by a sample of surveyed participating teachers to be aligned with principles of high-quality PTELs described in the Quality of Project Design section of this proposal. |
| Develop scoring strategies | Procedures and rubrics for rating student responses to PTELs and Mini PTs are judged by Advisory Board as suitable for operational use based on field test data.  Procedures for combining information from Common PTEL and End-of-Year assessment components are judged by Advisory Board as suitable for operational use based on field test data.  Automated scoring procedures are judged by Advisory Board as being properly evaluated psychometrically and worthy of continued research and development based on field test data. |
| Iteratively analyze the performance tasks | Functioning of PTELs and Mini PTs is judged by Advisory Board members to have been competently analyzed through cognitive labs, pilot testing, and field testing, and appropriately revised after each data collection.  A majority of educator interviewees judged that the PTELs and Mini PTs were designed with a clear instructional purpose, comprehensible, NGSS-aligned, and grade-level appropriate, and could feasibly be implemented in a classroom setting.  A majority of educators and students judged that the PTELs and Mini PTs were useful for informing instruction. |
| Provide professional development services to educators | A majority of educators judged the PD experience useful for understanding how to score the PTELs and Mini PTs consistently and accurately.  A majority of educators judged PD activities enhanced their general knowledge of assessment practices in science and their specific knowledge of using formative assessment practices to promote student learning.  A majority of educators judged PD activities aligned with the practices of high-quality teacher PD described in the Project Design section of this proposal. |

## References

AP Central. (n.d.-a). *AP computer science principles*. College Board. <https://apcentral.collegeboard.org/courses/ap-computer-science-principles/course>

AP Central. (n.d.-b). *AP research*. College Board. <https://apstudents.collegeboard.org/courses/ap-research/assessment>

AP Central. (n.d.-c). *AP seminar.* College Board. <https://apstudents.collegeboard.org/courses/ap-seminar>

Badrinarayan, A., & Darling-Hammond, L. (in press). Design principles for instructionally relevant assessment systems. *Learning Policy Institute.*

Bang, M., & Vossoughi, S. (2016). Participatory design research and educational justice: Studying learning and relations within social change making. *Cognition and instruction*, *34*(3), 173-193. <https://doi.org/10.1080/07370008.2016.1181879>

Bennett, R. E. (2023). Toward a theory of socioculturally responsive assessment. *Educational Assessment*, *28*(2), 83–104. <https://doi.org/10.1080/10627197.2023.2202312>

California Department of Education. (2017). *California Science Test blueprint.*

California Department of Education. (2020). *Revised California Science Test blueprint.*

California Department of Education. (2022). *NGSS for California public schools, K-12*. <https://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp>

California Department of Education. (2023). *Science*. <https://www.cde.ca.gov/pd/ca/sc/>

ClimeTime Assessment Project. (n.d.). *Assessment resources*. <https://www.climetime.org/assessments/resources/>

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37–46. <https://doi.org/10.1177/001316446002000104>

Colorado Department of Education. (2023). *Assessment resource bank.* <https://www.cde.state.co.us/assessment/resourcebank-assessments>

Connecticut State Board of Education. (2003) Connecticut Academic Performance Test [Discontinued]. <https://www.csde.state.ct.us/public/der/s-t/testing/capt/capt_program_overview_web.pdf>

Contextus.Science. (n.d.). Home: <https://contextus.science/>

Cumbo, B., & Selwyn, N. (2022). Using participatory design approaches in educational research. *International Journal of Research & Method in Education*, *45*(1), 60–72. <https://doi.org/10.1080/1743727X.2021.1902981>

Darling-Hammond, L., Burns, D., Campbell, C., Goodwin, A. L., Hammerness, K., Low, E. L., ... & Zeichner, K. (2017). *Empowered educators: How high-performing systems shape teaching quality around the world*. John Wiley & Sons.

Darling-Hammond, L. (1996). Performance-based assessment and educational equity. In E. R. Hollins (Ed.), *Transforming curriculum for a culturally diverse society* (pp. 245–272). Routledge. <https://doi.org/10.4324/9781315045634-16>

Darling-Hammond, L. (2017). Developing and measuring higher order skills: Models for state performance assessment systems. Council of Chief State School Officers. <https://files.eric.ed.gov/fulltext/ED606777.pdf>

ETS. (2021). *Best practices for constructed-response scoring.* <https://www.ets.org/pdfs/about/cr_best_practices.pdf>

Evans, C. M. (2019). Effects of New Hampshire’s innovative assessment and accountability system on student achievement outcomes after three years. *Education Policy Analysis Archives*, *27*(10), n10.

Fine, C. G., & Furtak, E. M. (2020). A framework for science classroom assessment task design for emergent bilingual learners. *Science Education*, *104*(3), 393–420. <https://doi.org/10.1002/sce.21565>

Fine, C. G., & Furtak, E. M. (2024). Asset and deficit discourses of student ideas in science formative assessment co‐design. *Journal of Research in Science Teaching*. <https://doi.org/10.1002/tea.21944>

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. Psychological Bulletin, 76(5), 378–382. <https://doi.org/10.1037/h0031619>

Goodman, L. A., & Kruskal, W. H. (1954). Measures of association for cross classifications. *Journal of the American Statistical Association*, *49*, 732–769. <https://doi.org/10.2307/2281536>

Heritage, M., & Wylie, E. C. (2020). Formative assessment in the disciplines: Framing a continuum of professional learning. Harvard Education Press.

Liu, L., Cisterna, D., Kinsey, D., Qi, Y., & Steimel, K (in press). AI-based diagnosis of student reasoning patterns in NGSS assessments. In X. Zhai & J. Krajcik (Ed.), *Uses of Artificial Intelligence in STEM Education*. Oxford University Press.

Liu, L., Rogat, A., & Bertling, M. (2013). A CBAL™ science model of cognition: Developing a competency model and learning progressions to support assessment development. *ETS Research Report Series*, *2013*(2), 1–54. <https://doi.org/10.1002/j.2333-8504.2013.tb02336.x>

Luck, R. (2018). What is it that makes participation in design participatory design? *Design studies*, *59*, 1–8.

National Academies of Sciences, Engineering, and Medicine. (2019)*. Monitoring educational equity*. National Academies Press.

National Assessment Governing Board (2024)*. The 2028 NAEP science assessment framework*. <https://www.nagb.gov/naep-subject-areas/science/2028-naep-science-assessment-framework.html>

National Assessment of Educational Progress. (2018). *2014 and 2018 Technology and Engineering Literacy (TEL) Assessments*. US Department of Education, Institute of Education Sciences, National Center for Education Statistics. <https://www.nationsreportcard.gov/tel/tasks/>

National Assessment of Educational Progress. (2019). *2019 Science Assessment*. US Department of Education, Institute of Education Sciences, National Center for Education Statistics. <https://www.nationsreportcard.gov/science/sample-questions/?grade=4>

New Hampshire Department of Education. (2023). *Performance Assessment of Competency Education* *(PACE)*. Retrieved 14 September 2023 from <https://nhdoe.instructure.com/courses/58/modules#module_87>

[NGSS Lead States. (2013](https://www.nextgenscience.org/)). *Next generation science standards: For states, by state*s. National Academies Press. <https://www.nextgenscience.org/>

Pellegrino, J. W., Gane, B. D., Zaidi, S. Z., Harris, C. J., McElhaney, K. W., Alozie, N., ... & Van Horne, K. (2018). The challenge of assessing “knowledge in use”: Examples from three-dimensional science learning and instruction. *International Society of the Learning Sciences, Inc*.[ISLS].

Penuel, W. R., Riedy, R., Barber, M. S., Peurach, D. J., LeBouef, W. A., & Clark, T. (2020). Principles of collaborative education research with stakeholders: Toward requirements for a new research and development infrastructure. *Review of Educational Research*, *90*(5), 627–674. <https://doi.org/10.3102/0034654320938126>

Performance Assessment Resource Bank. (n.d.). *Performance Assessment Resource Bank*. <https://www.performanceassessmentresourcebank.org/>

Queensland Curriculum & Assessment Authority. (n.d.). *Years 9–10 assessment techniques and conditions.* Queensland Government. <https://www.qcaa.qld.edu.au/downloads/aciq/science/assessment/ac_science_yr9-10_tc.pdf>

Shepard, L. A., Diaz-Bilello, E. K., Penuel, W. R., & Marion, S. F. (2020). Classroom assessment principles to support teaching and learning. *Center for Assessment, Design, Research and Evaluation, University of Colorado Boulder*. <https://www.colorado.edu/cadre/sites/default/files/attached-files/classroom_assessment_principles_to_support_teaching_and_learning_-_final_0.pdf>

Singapore Examinations and Assessment Board. (2023). *O-level syllabuses examined in 2023*. Cambridge Assessment International Education. <https://www.seab.gov.sg/home/examinations/gce-o-level/o-level-syllabuses-examined-for-school-candidates-2023>

Stackable, Instructionally-Embedded, Portable Science Partners. (2021). *SIPS resources*. <https://sipsassessments.org/resources/>

Stecher, B. (2010). Performance assessment in an era of standards-based educational accountability. *Stanford Center for Opportunity Policy in Education*.

Web Content Accessibility Guidelines 2.0, W3C World Wide Web Consortium Recommendation (http://www.w3.org/TR/200X/REC-WCAG20-20081211/, Latest version at <http://www.w3.org/TR/WCAG20/>).

1. Source of the data is from <https://www.cde.ca.gov/ds/ad/ceffingertipfacts.asp>. [↑](#footnote-ref-2)