

California Department of Education

Executive Office

SBE-003 (REV. 11/2017)

imab-adad-jan24item03

# California State Board of Education January 2024 Agenda Item #03

## Subject

California Assessment of Student Performance and Progress: Approval of the Receipt of the California Science Test Innovations Concept Paper.

## Type of Action

Action, Information

## Summary of the Issue(s)

The California Department of Education (CDE) seeks approval of the receipt of the California Science Test (CAST) Innovations Concept Paper as part of the California Assessment of Student Performance and Progress (CAASPP). This concept paper describes the research conducted by the contractor and a proposal for an innovative approach for future design and development of the CAST. The proposed approach would include performance tasks embedded in learning (PTELs) that provide teachers and students authentic and engaging experiences through doing science.

## Recommendation

The CDE recommends that the California State Board of Education (SBE) approve 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 CDE also requests feedback about the concepts outlined in the paper and direction from the SBE on next steps.

## Brief History of Key Issues

The following is a summary of the CAST Innovations Concept Paper.

### History of the California Science Test

In November 2013 the SBE adopted the *Next Generation Science Standards for California Public Schools, Kindergarten Through Grade Twelve (K–12)*. The California Next Generation Science Standards (CA NGSS) emphasize three dimensions of science learning, which consist of the following:

* Disciplinary Core Ideas (DCIs) are at the core of four disciplines of science: Earth and Space Sciences, Life Sciences, Physical Sciences, and Engineering.
* Crosscutting Concepts (CCCs) apply across multiple disciplines of science and engineering.
* Science and Engineering Practices (SEPs) are real-world behaviors and processes used by scientists and engineers.

For more information, please visit the CDE NGSS for California Public Schools, K–12 web page at <https://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp>.

The CAST was created to assess the CA NGSS. The CDE and ETS consulted with a large and diverse group of experts to gather feedback on priorities for the CAST. This group included:

* Hundreds of California science teachers, including representatives of the California Association of Science Educators, formerly known as the California Science Teachers Association;
* Higher education officials, including Helen Quinn as a representative of Stanford’s NGSS Assessment Program;
* Science, technology, engineering, and mathematics experts, including Jim Pellegrino as a representative from the National Academy of Science’s National Research Council Framework for K–12 Science Education; and
* Representatives of various assessment interest holders and advocacy organizations

The goals of the CAST included the following:

* Promote improvements to teaching and learning.
* Incentivize science instruction in every grade level.
* Measure the range and depth of the CA NGSS.
* Provide models of high-quality items that reflect fidelity to the CA NGSS.
* Minimize testing time and costs while providing accessibility for all students.

The SBE approved the CAST high-level test design in March 2016 and the blueprint in November 2017.

The CAST was administered as a census pilot test in 2016–17, a census field test in 2017–18, and an operational test in 2018–19. After reviewing data from the first operational year, a decision was made to revise the original CAST blueprint to best represent all science domains and reduce overall testing time.

In January 2020, the SBE approved the revised CAST blueprint, which consists of the following:

* Segment A contains discrete items that assess the breadth of content across all three CA NGSS domains.
* Segment B contains three sets of items that add to the depth of domain coverage. Each set covers one science domain using a set of items of increasing complexity, including a constructed-response item.
* Segment C contains field test items.

In December 2021, the CAST was submitted to the U.S. Department of Education (ED) for assessment peer review. The federal peer review process evaluates the technical quality of an assessment and ensures its consistency with statutory and regulatory requirements outlined in the federal Every Student Succeeds Act. The CAST substantially met all requirements for federal peer review on the initial submission (which included the original blueprint [2018] and the revised blueprint [2020]), and the CDE submitted additional documentation to fulfill the outstanding requirements in June 2023. The CDE anticipates receiving feedback from the ED in early 2024.

July 2022 marked the beginning of a new California Assessment Systems contract with ETS. Task 6.8 is devoted to exploring further assessment innovations for the CAST.

For Task 6.8, two goals were specified:

1. Develop PTELs for the CAST
2. Further revise the CAST blueprint to shorten testing time for the end-of-year assessment.

Common definitions of PTs assume that students produce an analysis, product, performance, or solution to a creative or investigative task that might take a class period or less, up to several weeks or months.

### An Innovative Conceptual Approach for Science Assessment

Since 2022, the CDE worked with assessment contractor ETS to conduct a review of existing performance tasks (PTs) and assemble a framework for the use of PTs in assessment systems. The review identified a set of features that can be clustered into three broad categories: NGSS-aligned, student-centered, and instructionally relevant. While each feature contributes to the design of high-quality PTs, integrating all features into a single PTEL may not always be feasible. Therefore, a group of science educators and assessment experts identified those that they recommend should be prioritized for implementation in the contexts of classroom and summative assessment use. The prioritized features noted that assessments should accomplish the following:

* **Be authentic:** Assessments highlight and center the key concepts, modes of inquiry, and ways of learning in the discipline.
* **Represent the NGSS dimensions:** Tasks facilitate students’ application and integration of the DCIs, SEPs, and CCCs.
* **Involve sensemaking of phenomena:** Tasks create opportunities to meaningfully engage with scientific phenomena by placing phenomena within contexts likely to be familiar to students.
* **Reflect and be responsive to learners:** Assessments 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.
* **Are useful for informing decisions that impact instruction:** Assessments help encourage, model, and inform high-quality curriculum and instruction.

The proposed conceptual approach to address these priorities comprises four assessment-related system components:

1. **Through-year classroom PTELs,** readily available for teachers in all grade levels as tasks that support formative assessment and instruction of the CA NGSS
2. **Through-year common PTEL(s)**to supplement the end-of-year summative assessments in grade five, grade eight, and high school, both contributing to measurement and positively influencing instruction
3. **End-of-year** **summative/state assessment that includes a mini PT**
4. **Opportunity-to-learn in science indicators**to help place test scores in context

Implications for implementation across the variations in models of science education will require careful consideration and resolution. The specific timeline for any possible future actions is yet to be determined.

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

In November 2021, the CDE provided the SBE with contract number CN220002 with ETS for the administration of the CAASPP and the English Language Proficiency Assessments for California, which includes the concept paper for incorporating assessment innovations into the California Assessment System (<https://www.cde.ca.gov/be/ag/ag/yr21/documents/nov21item04.docx>).

In a December 2019 Information Memorandum, the CDE provided the SBE with updates on the results of ETS’ CAST studies and analyses on the multistage test, the screener, the use of matrix sampling for group reporting, and the amount of time it takes students to complete the CAST (<https://www.cde.ca.gov/be/pn/im/documents/dec19memoadad01.docx> and <https://www.cde.ca.gov/be/pn/im/documents/dec19memoadad01a01.docx>).

In a December 2019 Information Memorandum, the CDE provided the SBE with a summary of the *CAASPP 2019 Independent Evaluation Report* and notified the SBE of the posting of the full report on the CDE website (<https://www.cde.ca.gov/be/pn/im/documents/dec19memoadad03.docx>).

In October 2019, the CDE provided the SBE with updates on the CAST Standard Setting Plan and a summary of related activities conducted before and after the convening of the CAST July/August 2019 Standard Setting Workshop (<https://www.cde.ca.gov/be/pn/im/documents/oct19memoadad01.docx>).

In August 2019, the CDE provided the SBE with updates on the upcoming development activities and test format for the 2019–20 CAST administration (<https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-adad-aug19item01.docx>).

In December 2018, the CDE provided the SBE with an Information Memorandum that gave an update on the 2017–18 public releases of the preliminary indicators for the CAST field test (<https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-adad-dec18item01.docx>).

In August 2018, the CDE provided the SBE with an Information Memorandum that gave an update on the development of the new science assessment, the CAST (<https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-adad-aug18item01.docx>).

In May 2018, the CDE provided the SBE with updates on the CAASPP System including the full-census field test, CAST Academy training, Preliminary Indicator scoring toolkit (<https://www.cde.ca.gov/be/ag/ag/yr18/documents/may18item03.docx>).

In March 2018, the CDE provided the SBE with updates on the CAASPP System including the development of resources and training to support the CAST field test (<https://www.cde.ca.gov/be/ag/ag/yr18/documents/mar18item08.docx>).

The SBE adopted the CA NGSS in September 2013. The Implementation Planwas adopted by the SBE in November 2014. The CDE provided information memoranda regarding CA NGSS implementation to the SBE in December 2015, June 2016, September 2017, and February 2018.

## Fiscal Analysis (as appropriate)

There is potential future fiscal impact tied to this program. The CDE will be able to determine any potential future cost once the concept paper is approved and the SBE has provided feedback and direction.

## Attachment(s)

* Attachment 1: California Science Test Innovations Concept Paper (39 pages)



**California Science Test Innovations Concept Paper**

**Contract #CN220002**

**Prepared for the California Department of Education by ETS**

**Presented December 8, 2023**



**Contents**

[1. Executive Summary 4](#_Toc152935606)

[2. Review of Science Performance Tasks 8](#_Toc152935607)

[3. A New Conceptual Approach for an Assessment System 10](#_Toc152935608)

[4. The Theory of Action 18](#_Toc152935609)

[5. Scenarios for Phased Implementation of the Proposed Conceptual Approach 20](#_Toc152935610)

[6. Future Directions and New Horizons 30](#_Toc152935611)

[7. References 31](#_Toc152935612)

[8. Accessibility Information 33](#_Toc152935613)

[9. Appendix: Development of the Proposed Conceptual Approach 35](#_Toc152935614)

**List of Figures**

[Figure 3.1 Schematic of conceptual approach 11](#_Toc152739359)

[Figure 4.1 Theory of action for the conceptual approach 19](#_Toc152739360)

[Figure 5.1 Simultaneous phasing scenario 21](#_Toc152739361)

[Figure 5.2 Sequential phasing scenario beginning with classroom PTELs 24](#_Toc152739362)

[Figure 5.3 Sequential phasing scenario beginning with modifications to the end-of-year assessment 26](#_Toc152739363)

[Figure 5.4 Parallel development of the PTELs 28](#_Toc152739364)

**1. Executive Summary**

The purpose of this paper is to present a conceptual approach, including a theory of action, for California Science Test (CAST) innovation. The description of the conceptual approach is complemented by a brief summary of a review of relevant innovative assessments, as well as insights gained from educator learning sessions.

**Introduction and Background**

The CAST was created to assess the [*Next Generation Science Standards for California Public Schools, Kindergarten Through Grade Twelve*](https://www.cde.ca.gov/pd/ca/sc/ngssstandards.asp)*,* which were adopted by the California State Board of Education (SBE) in November 2013 (California Department of Education [CDE], 2022). The California Next Generation Science Standards (CA NGSS) emphasize three 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. SEPs are real-world practices engaged by practicing scientists and engineers.

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 kindergarten through grade twelve science education equips students through learning and engagement 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).

**Development Process**

After the adoption of the standards, the CDE and ETS began work with science educators, national experts, and interest holder groups to determine what the CAST should include. The first meeting was held in July 2014. Educators provided input to inform the decision for the CAST to be administered to all students in grades five and eight and one time in high school (i.e., grade ten, eleven, or twelve), unless a student’s individualized education program (IEP) designates the California Alternate Assessment. Because the CDE recommends testing high school students when they are enrolled in their last science course, high schools have the option to test any or all students in grade ten or eleven as long as all students have been tested by the end of grade twelve. Additionally, educators emphasized the importance of the specific grade-level tests being inclusive of science content from previous grade levels. This feedback helped inform the next steps for the CAST.

Initial goals of the CAST included the following:

* Promote improvements to teaching and learning
* Incentivize science instruction in every grade level
* Measure the range and depth of the CA NGSS
* Provide models of high-quality items that reflect fidelity to the CA NGSS
* Minimize testing time and costs while providing accessibility for all students

ETS produced a white paper on using evidence-centered design as the basis for developing the CAST (ETS, 2019). The high-level test design was reviewed by science educators and national experts, including James Pellegrino, Randy Bennett, and several Stanford NGSS Assessment Project contributors (including Helen Quinn and others).

The initial test design was intended to include three segments:

* Segment A would be a two-stage adaptive section that presented items intended to be substantively equivalent in content, but which may differ in difficulty to match each student’s estimated level of competency.
* Segment B would contain two tasks designated as PTs, here indicated with an asterisk as “PTs”\* because their design differed from commonly accepted definitions of such tasks. Common definitions assume that students produce an analysis, product, performance, or solution to a creative or investigative task that might take a class period or less, up to several weeks or months. The CAST PTs,\* in contrast, were intended to add depth-of-domain coverage, minimize testing time, and be efficient to score. As such, they would contain four to six short selected-response items each, set in the same content-domain context.
* The content presented in this PT\* segment would be adapted to the student’s performance in Segment A. Where a student demonstrates especially low performance in a particular domain in Segment A, the student would not be presented with items based on the same domain, which would yield little or no additional information. In this way, Segment A serves to screen out content assigned to a student in Segment B. Where a student performs at similar levels in all science domains in Segment A, the assignment of items presented in Segment B would be random. Most Segment B assignments would follow this path.
* Segment C would contain either 13 discrete items or one CAST PT.\* Segment C would be intended to support embedded field testing of new items or the administration of operational items that could contribute additional, unique information at the group level via matrix sampling.

Items in segments A and B would contribute to reporting individual student scores. Items in all three segments—exclusive of field test items—were intended for reporting group-level scores. Group-level reporting was to be defined on the basis of further input from science educators and national experts.

This high-level test design was approved by the SBE in March 2016 (ETS, 2016). From there, a blueprint was developed to ensure equitable content coverage to support both an overall science score and a performance level indicator for each of the three content domains (Earth and Space Sciences, Life Sciences, and Physical Sciences) (CDE, 2017). The blueprint was approved in November 2017.

The CAST was administered as a census pilot test in 2016–17, a census field test in   
2017–18, and operationally in 2018–19. Data from the first operational year was gathered to inform the practicality and utility of several features of the high-level design. Data from additional studies resulted in a decision to

* eliminate the two-stage adaptive design and reevaluate its utility once the CAST item bank has been built to include a wider range of items at the higher difficulty levels;
* eliminate the use of the content screener for PTs;\*
* add a third PT\* so that all three science content domains would be represented; and
* reduce the number of discrete items to lower overall testing time.

These decisions were approved by the SBE in January 2020 as part of a revised CAST blueprint (CDE, 2020). In December 2021, the CAST was submitted to the U.S. Department of Education for assessment peer review. The federal peer review process evaluates the technical quality of an assessment and ensures its consistency with statutory and regulatory requirements outlined in the federal Every Student Succeeds Act. The CAST substantially met all requirements for federal peer review on the initial submission (which included the original blueprint [2018] and the revised blueprint [2020]). The CDE submitted additional documentation to fulfill the outstanding requirements in June 2023.

**The Next Iteration**

To fully realize the vision of the CA NGSS, the CDE and the SBE have requested that ETS explore innovative science PTs that support science learning and instruction by providing teachers and students with authentic and engaging experiences that *measure* science through *doing* science. To assist teachers in the implementation of the CA NGSS and to support diverse students’ learning needs, the CDE and SBE are conceptualizing an innovative assessment system that focuses on PTs that can be embedded in learning (also known as PTEL. A PTEL would allow students to *perform* to demonstrate their knowledge and conceptual understanding.

July 2022 marked the beginning of a new California Assessment Systems contract with ETS. Two sections of the contract included new work related to the science assessment:

1. Task 6.1 includes the development of CAST Interim Assessments.
2. Task 6.8 is devoted to exploring further assessment innovations for the CAST.

The interim assessments are intended to give teachers timely feedback as to what their students know and can do with respect to the CA NGSS so that instruction can be adjusted, as appropriate. Additionally, the CDE has developed—and continues to develop—[Tools for Teachers](https://smartertoolsforteachers.org/), which are resources that could be used independently or as a complement to the new interim assessments.

For Task 6.8, two goals were specified:

1. Developing PTELs for the CAST
2. Further revising the CAST blueprint to shorten the time students spend on the end-of-year assessment

Additional goals may be identified by the CDE and SBE.

**2.** **Review of Science Performance Tasks**

Since 2022, the CDE and SBE have worked with ETS to review existing PTs and assemble a framework for the use of PTs in assessment systems. The review work commenced by conducting learning sessions with national experts and practitioners to understand their experiences developing and using PTs. The project team engaged in two learning sessions with eight California science educators and nine national experts. Next, the team analyzed 39 PT initiatives to develop a comprehensive list of features and then clustered these features into three categories essential to high-quality PTELs. The review identified a set of features that can be clustered into three broad categories:

1. NGSS Aligned, emphasizing consistency with the NGSS
2. Student Centered, highlighting the need to center tasks around students’ experiences, prior knowledge, and how they learn
3. Instructionally Relevant, providing both models and information for instruction

Each of the features fit into at least one category, with some features belonging to two categories.

The work culminated in four potential conceptual approaches (refer to the [appendix](#_Appendix:_Development_of)) to implementing PTELs and leveraging their capabilities to better inform instruction and assessment. Through this process, the team arrived at a conceptual approach consistent with the vision of the CA NGSS.

**To Support Instructionally Aligned Assessment**

To understand what features should be prioritized, the team compared the high-quality PTEL features that were identified with those described in other investigations of PT quality. Documents were reviewed, such as the Science Assessment for Emergent Bilingual Learners checklist (SAEBL; Fine & Furtak, 2020), Science Task Prescreen tool (Achieve, 2018), and the Design Principles for Instructionally Relevant Assessment Systems (Badrinaryan & Darling-Hammond, in press). The team found that each category of PTEL features was also included in at least one of these other documents.

While each category of the features from the combined list presented previously contributes to the design of high-quality PTs, integrating all features into a single PTEL may not always be feasible. As such, ETS subsequently requested a group of science education and assessment experts to review the individual features identified from the reviewed documents and select those that should be a priority for implementation in the contexts of classroom and summative use. The features that were most frequently noted as priorities were that assessments meet the following criteria:

* **Be authentic:** Assessments highlight and center the key concepts, modes of inquiry, and ways of learning in the discipline.
* **Represent the NGSS dimensions:**Tasks facilitate students’ application and integration of the DCIs, SEPs, and CCCs.
* **Involve sensemaking of phenomena:** Tasks create opportunities to meaningfully engage with scientific phenomena by placing phenomena within contexts likely to be familiar to students.
* **Reflect and are responsive to learners:** Assessments 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.
* **Are useful for informing decisions that impact instruction:** Assessments help encourage, model, and inform high-quality curriculum and instruction.

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

* provide authentic opportunities for students to engage in CA NGSS—defined dimensions of science learning and receive feedback or related instruction from teachers throughout the year;
* help teachers gain experience in understanding how to support NGSS-aligned science inquiry as well as how to assess evidence of student proficiency; and
* become readily available through 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. Furthermore, PTELs have the potential to play a role in the reimagining of the state summative assessment, as described next.

**3.** **A New Conceptual Approach for an Assessment System**

The review presented previously undergirds a conceptual approach for an innovative assessment system that addresses the identified priorities. The proposed conceptual approach comprises four assessment-related system components (figure 3.1):

1. Through-year classroom PTELs, readily available for teachers in all grade levels as tasks that support formative assessment and instruction of the CA NGSS
2. Through-year common PTEL(s) to supplement the end-of-year summative assessment in grade five, grade eight, and high school, both contributing to measurement and positively influencing instruction
3. End-of-year summative/state assessment that includes a mini PT
4. Opportunity-to-learn in science indicators to help place test scores in context

These four components, which are each described in more detail in the next subsection, form a cohesive system focused on serving instruction and state summative assessment purposes. Each through-year PTEL, whether a classroom or common PTEL, is constructed upon the same CA NGSS foundation. Synergy between the common and classroom PTELs that could be engineered through overlapping contexts and data sets may foster better integration of teaching, learning, and assessment. By design, the through-year classroom PTELs also provide teachers the flexibility to adapt materials to best support students’ sensemaking of phenomena by providing personalized and student-centered learning opportunities.

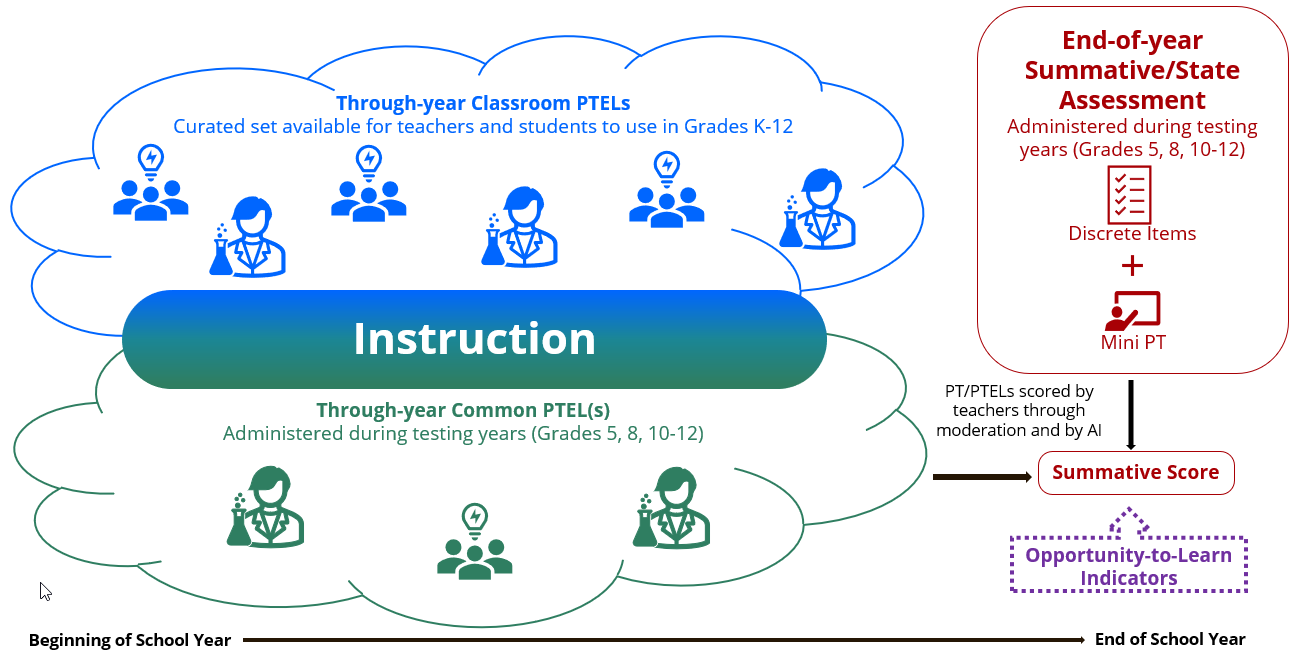
The components of the proposed summative portion of the assessment system (through-year common PTELs and end-of-year summative assessment) will be tailored to, and administered solely in, grade five, grade eight, and once in grade ten, eleven, or twelve. However, the curated set of classroom PTELs would be available to teachers in all grade levels as an aid to instructional and classroom assessment practices.

The structure of science education in high school tends to be different from that in middle school, which in turn differs from the organization of science education in elementary school. Notable differences include the following:

* In high school, it is the presence of discipline-specific courses, with some schools opting for integrated science.
* In middle school, there is a broader mix of general or integrated and discipline-specific courses.
* Elementary schools typically have science taught by generalist teachers.

The implications of these differences for the details of design and implementation fall outside the scope of this report. However, they will require careful consideration and resolution should the conceptual approach be approved.

Figure 3.1 presents a schematic representation of how these four components are arranged into an assessment system. Note that the number of common PTELs is yet to be determined.



**Figure 3.1 Schematic of conceptual approach**

**Through-Year Classroom Performance Tasks**

The first component is a curated set of PTELs intended for classroom use to support instruction and formative assessment at the teacher’s discretion any time during the year. These 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 form 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’ individual 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 teachers may generate.

**Models for the Classroom Performance Tasks**

There are several examples of existing resources, any of which might serve as a model for developing the curated set of classroom PTELs or as a source for licensing content.

* [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, PTs, 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 PTs. 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 involving Nebraska, New Mexico, and Montana. It aims to develop a collection of resources, including PTs and instructional materials. Nebraska, for example, has developed a Science Classroom Formative Task Repository. This repository includes PTs 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.

In each of these examples, task collections are intended to make it easier for teachers to find, adapt, and integrate high-quality tasks into their instructional and assessment practice. The tasks are typically organized by subject area, grade level, and standard, making it convenient for educators to align assessments with their curriculum and learning objectives.

**Through-Year Common Performance Tasks**

The second component is a curated set of PTELs, which, at the state’s discretion, could be used in a through-year assessment fashion within the state testing window to supplement the end-of-year CAST (i.e., contribute to the overall summative score). To distinguish this component from the classroom PTELs, these are referred to as “common PTEL(s)” henceforth. These common PTELs would represent more standardized versions of the classroom PTELs, intended for administration without major modifications, unless such modifications were approved, for example, to accommodate English learner students or students with IEPs.

The common PTELs will closely align with the CA NGSS performance expectations for the testing grade band, thereby making possible the combination of results with students’ end-of-year summative score(s). To address issues of variation in scope and sequence across classrooms, schools, and LEAs, the order and timing of test administration within the designated year could be determined at the discretion of the LEA, the school (if the LEA delegates that decision), or the teacher (if the school delegates that decision).

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 1 to 2 or 3, at the option of the state). The purpose of the common PTELs is to, first and foremost, encourage focus on applications of the CA NGSS SEPs. For each occasion, design possibilities could include administering the same PTEL to all students in a grade level, spiraling multiple parallel versions across students, classes, or schools to increase reproducibility; or spiraling PTELs that differ in the SEP, DCI, or CCC to maximize blueprint coverage. Decisions would need to be made regarding the domain that the PTELs will focus on within a given year and grade level so that balance of item coverage across domains is taken into account.

For students in grades ten through twelve, the common PTELs are envisioned to be less-scaffolded, inquiry-based tasks than in the earlier grade levels, resembling tasks requiring student agency like those in the Advanced Placement® (AP) Computer Science Principles examination, which is described in more detail in the next subsection.

The common PTEL(s) can be designed in various formats and durations of implementation, depending upon the state’s preferences, including projects, shorter technology-supported tasks, hands-on activities with documentation of student work captured digitally and uploaded, or combinations of those forms. Additionally, the common PTEL(s) could involve collaborative portions, with the possibility of scoring the contributions of the group, individuals, or both.

**Models for the Common Performance Task(s)**

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 PT prototypes
* New Hampshire’s Performance Assessment of Competency Education (PACE) program common tasks
* Washington State’s ClimeTime tasks
* Certain AP examinations such as Computer Science Principles (AP Central, n.d.-‍a), Seminar (AP Central, n.d.-c), and Research (AP Central, n.d.-b)
* Additional PTs used for accountability purposes in Australia and Singapore

Several formerly administered science assessments, such as those of Connecticut and Vermont, may also provide examples of common PTELs and are as follows:

The CBAL(Liu et al., 2013) science PT prototypes were designed to assess student understanding in through-course fashion, rather than relying solely on measurement at a single point. This PT prototype was developed for formative assessment purposes and designed to be administered during a two- to three-week instructional period.

* [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 district’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 provides teacher guides, task facilitation slides, student task documents, and sensemaking tools for educators and learners, which could serve as a model for the supports and resources required for some level-setting of 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 PTs. However, some examinations 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). To varying degrees, these examinations each require students to demonstrate their understanding and skills in examination 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 culminating presentation could constitute both the learning experience and its assessment.
* Several international assessment systems in science include components similar to the common PTEL described previously. In Queensland, Australia, science PTs 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 student 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 (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.

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**

The third component is an end-of-year summative assessment. As in the current CAST, the assessment would include a section of discrete items. It would also include a mini PT, designed to align with the key features of the through-year classroom and common PTELs (e.g., shared data sources, shared contexts or phenomena, question formats, etc.), while being streamlined and logistically suitable for an end-of-year accountability test.

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), NAEP technology and engineering literacy assessment (NAEP, 2018), and in pilots being conducted by various states.

**Opportunity-to-Learn in Science Indicators**

The fourth component, opportunity-to-learn (OTL) in science indicators (hereafter referred to as OTL indicators), are indices intended to reflect the extent to which students have equal access to the resources needed to achieve the appropriate standards in 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. The audience for these indicators could be wide or narrow, depending upon the preferences of the state. Possibilities include LEA and school administrators, teachers, and parents/guardians, or selected subsets of these audiences. The inclusion of these indicators may also 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).

This assessment system component would be built upon existing state-collected data and, at the state’s option, supplemented by other sources. This data might include existing science-specific variables related to teacher preparation, teacher certification, science-specific professional development, science courses offered in high school (including advance levels), and availability of science enrichment programs.

Furthermore, valuable insights can be gained by including additional information from student, teacher, and school questionnaires administered along with the end-of-year assessment, at the state’s option. This information might include variables such as time devoted to science, curricular activities, availability of materials, and teachers’ opportunities for professional learning.

**Scoring Considerations**

The classroom and common PTELs, along with the mini PTs, will need to be scored in meaningful ways if the resulting information is to be useful. Fortunately, there is a long history with the scoring of PTs generally and many resources documenting that knowledge can be productively built upon by the state (e.g., ETS, 2021). The considerations are somewhat different for the several assessment system components—classroom PTELs, common PTELs, and end-of-year mini PTs—although there are points of overlap. Among the necessary decisions are ones related to who (or what) will do the scoring, what supports must be created to facilitate that work, and what monitoring processes might need to be put into place.

For the classroom PTELs, whether and how to score them will be a local decision. One option is for teachers to evaluate the responses of their students. Such an approach would require a general rubric along with the following resources for each PTEL, all of which could be part of the curated set of classroom PTELs:

* Topic notes indicating how the rubric applies to that PTEL
* Benchmark student responses associated with rubric descriptions and scores
* Exemplar student responses, both scored and unscored, to allow for teacher practice
* Teacher lesson plans for how to organize peer scoring among students and use it as part of the instructional process

Just as with teacher grading of any ambitious student work, a challenge with this approach is how to prevent scoring from becoming a burden that discourages use of the classroom PTELs. In some schools and LEAs, teacher collaboration or professional development time is allocated for scoring of PTs, along with discussion of curriculum and instructional implications from what is learned. As one possibility, teacher scoring of classroom PTELs could perhaps be supplemented with artificial intelligence (AI) methods, as such methods become more readily available.

The scoring of responses to common PTELs could be managed by a vendor as part of the assessment contract. AI methods could be employed to handle the majority of responses but supplemented by California teachers, who could perform back-reads as in the California Assessment of Student Performance and Progress (CAASPP) and other assessment programs. For this scoring, similar resources would need to be generated (i.e., general rubric, topic notes, benchmarks, range finders). Such teacher participation could be used as one means of improving knowledge of the CA NGSS and of student work, thereby enhancing the instructional capacity of participating teachers.

Scoring for the end-of-year assessment’s mini PTs could also be vendor managed and require the same types of rater resources as found in CAASPP.

For both the scoring of common PTELs and mini PTs, vendors would put into place monitoring procedures for observing and acting with respect to rater agreement, score drift within (and across assessment occasions, if more than one common PTEL occasion is used), and other potential error sources.

**4.** **The Theory of Action**

A theory of action is primarily intended to describe what a program is intended to achieve and how those goals are to be realized. The theory of action for the proposed conceptual approach depicts a holistic design aimed at influencing instruction and learning, as well as high-level decision-making (figure 4.1). Included in that design are the CA NGSS and professional support for teachers, as well as the four assessment components described previously: through-year classroom PTELs, through-year common PTELs, the end-of-year summative assessment that includes a mini PT, and OTL indicators.

The collective use of these elements is intended to achieve four primary goals (or ultimate outcomes):

1. Improved student learning and motivation in science
2. Enhanced teacher practices
3. More effective decisions by interest holders and policymakers
4. Increased community and public awareness regarding equitable learning opportunities and outcomes

Preceding these overarching goals are intermediate outcomes and the associated action mechanisms designed to bring about the intermediate and ultimate outcomes. 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 mini PTs, all action mechanisms. Engagement with these authentic PTs 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).

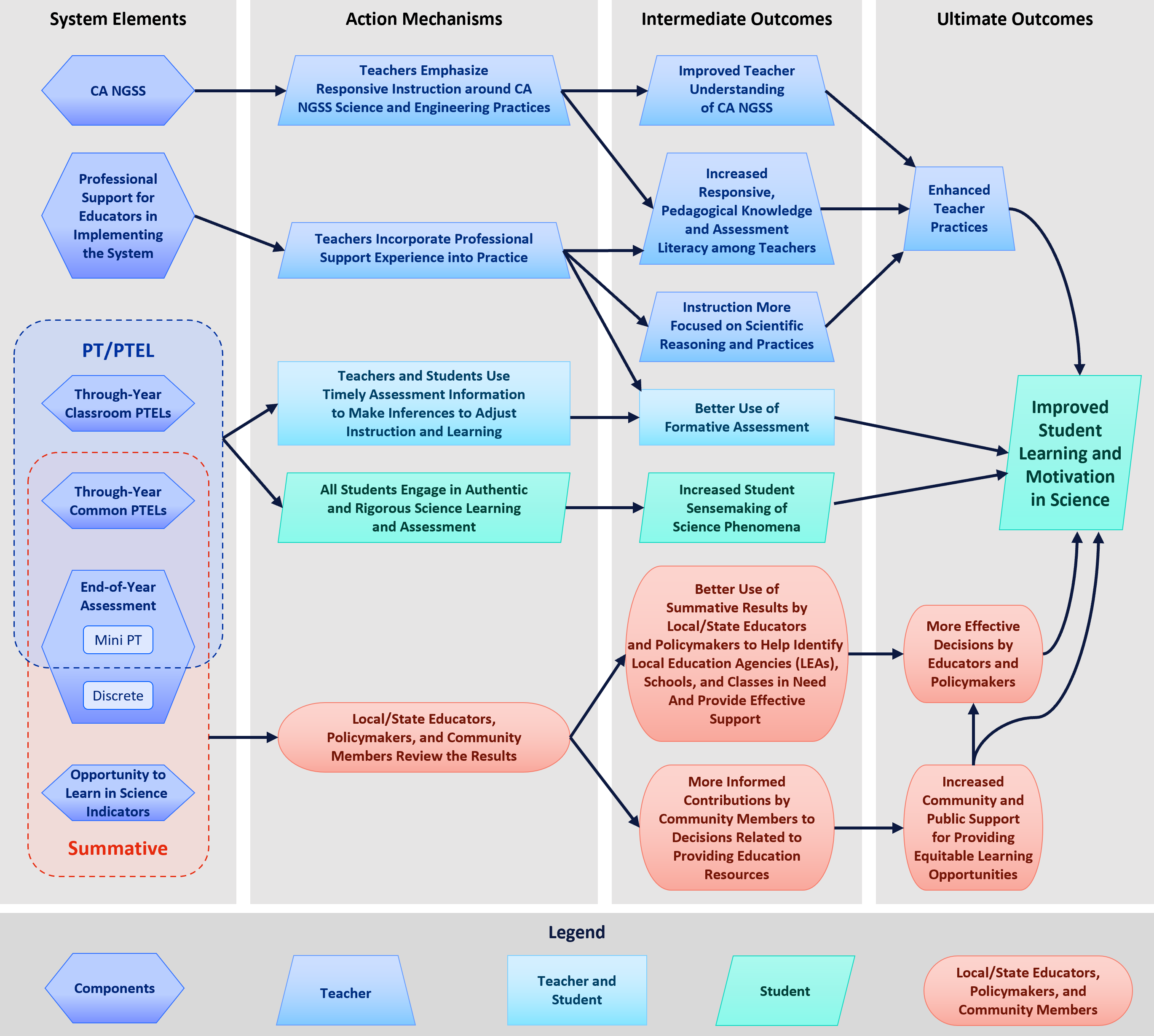
The end-of-year discrete items, end-of-year mini PTs, through-year common PTELs, and OTL indicators serve as tools for informing local or state policymakers and community members (action mechanism). Policymakers can, as appropriate, identify needs and provide necessary support to LEAs, schools, and classes, and community members contribute to decisions about the provision of education resources (intermediate outcomes).

Finally, teachers emphasizing instruction around the CA NGSS SEPs (action mechanism) should result in improved understanding of the CA NGSS, increased teacher responsive pedagogical knowledge, and greater assessment literacy (intermediate outcomes). These intermediate outcomes, in turn, should lead to the 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.

A flowchart for the theory of action for the conceptual approach is presented in figure 4.1. *Refer to the* [*Alternative Text for Figure 4.1*](#Alt_text_Figure_4_1) *for a description of this flowchart.*



**Figure 4.1 Theory of action for the conceptual approach**

**5.** **Scenarios for Phased Implementation of the Proposed Conceptual Approach**

Executing the conceptual approach for the assessment system will require meticulous engagement with various interest holders, such as teachers and school leaders. The specific timeline for implementation is yet to be determined. There are several potential scenarios for phased rollout of different components of the assessment system. The team envisions four general scenarios as plausible options.

In each scenario, for the classroom PTELs, pilot participation could be voluntary, subject to the discretion of local schools, LEAs, or teachers. For the common PTELs and end-of-year assessment components, it might be advisable to begin with a pilot involving a small sample of LEAs. The scores generated during the pilot phase would be considered experimental. The pilot would provide useful insights for the state to identify design and implementation challenges. An iterative process of piloting and fast design refinement might be implemented so that incremental improvements can be quickly implemented with minimal disruption for all involved.

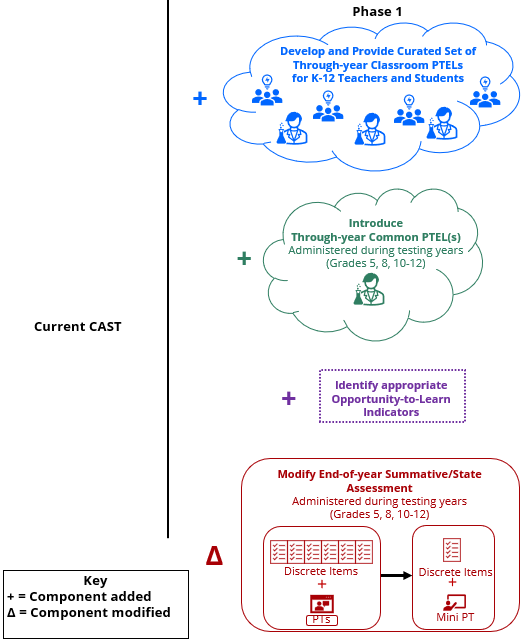
As design and implementation problems are identified and addressed, additional LEAs can be gradually added. This phased approach could unfold over several years with, for example, one-third of the state’s LEAs added in each of three years. Alternatively, more conservative phase-in distributions might also be considered (e.g., 10 percent of LEAs in the first year, 25 percent added in the second year, and the remaining 65 percent in the third year).

**Scenario 1: Simultaneous Phasing**

In this scenario, simultaneous implementation of several assessment system components is envisioned, including the curated set of classroom PTELs, the common through-year PTELs, and the modifications to the end-of-year summative assessment (figure 5.1).

A variation of this scenario could involve starting with the design and piloting of the modified end-of-year mini PTs alongside the design of the through-year common PTELs. However, the curated set of classroom PTELs could be under construction and put into use more gradually.

Figure 5.1 presents the single phase of the simultaneous phasing scenario.



**Figure 5.1 Simultaneous phasing scenario**

**Scenario 2: Sequential Phasing Beginning with Classroom Performance Tasks**

In this scenario, the phasing process occurs sequentially with at least three distinct phases (figure 5.2):

1. The implementation of a curated set of classroom PTELs
2. The development and implementation of the through-year common PTELs
3. The revision of the CAST

**Phase 1**

* Educators and assessment experts work together to codesign (or otherwise acquire) a curated set of classroom PTELs aligned with standards.
* OTL in science indicators are identified through a systematic review of information collected by the state; student and teacher questionnaires administered with state, national, and international assessments; science education literature; and other sources relevant to infrastructure and support systems that provide the conditions necessary for student success at LEA, school, and classroom levels.

**Phase 2**

* Common through-year PTELs are developed by the state or contractors.
* Once the curated set of classroom PTELs is provided, schools and LEAs can initiate administration of the common through-year PTEL(s) to students.
* As few as one through-year common PTEL may be administered. If additional through-year common PTELs will be administered, they could be implemented in the first year or phased in during subsequent years.

**Phase 3**

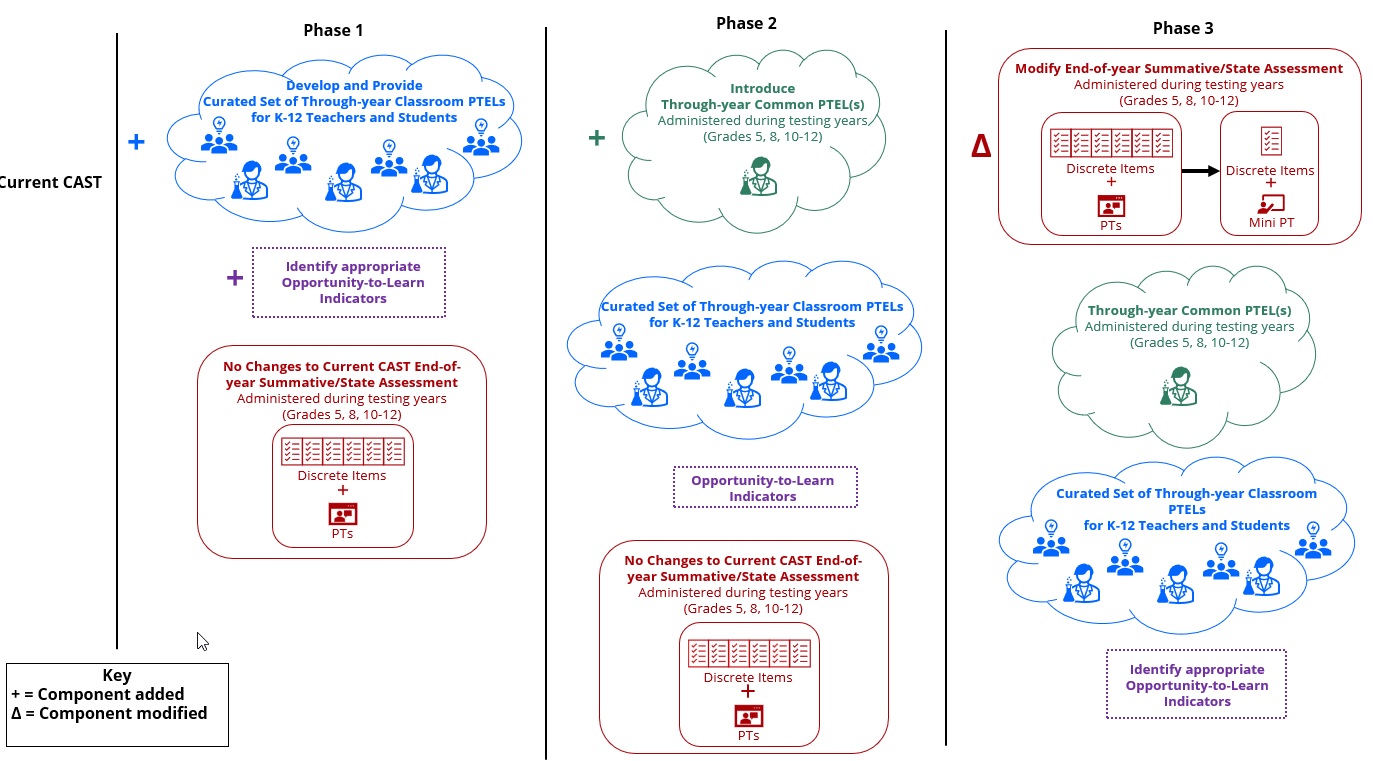
* In the final phase, the section containing discrete items on the current end-of-year summative assessment would be shortened, and a mini PT would be added.

**Scenario 2 Analysis**

This phased scenario offers several advantages. One advantage of this scenario is that changes to the CAST will occur within a condensed timeframe, which may be beneficial if educators are concerned about undergoing additional changes to the summative science assessments in subsequent phases during a long period of change. It also allows teachers time to use the curated set of classroom PTELs for formative and instructional purposes before changes to the CAST are introduced. It also provides time for adjustment to changes in the CAST before the implementation of the through-year common PTEL(s). Finally, the scenario would also allow for the field testing and refinement of classroom PTELs before developing the common PTEL(s) that would eventually contribute to a summative assessment score. Therefore, this scenario may afford greater opportunities compared to Scenario 1 for monitoring the processes of implementation of the classroom and common PTELs, including gathering teachers’ and students’ perceptions.

However, there is a potential downside in waiting to modify the end-of-year summative assessment, given it may be more challenging to integrate or create synergy between the common PTELs and the end-of-year mini-PT. Similarly, it might be more challenging to coordinate content coverage between the common PTELs and the discrete-item component of the end-of-year test. A clear disadvantage of this scenario is that the simultaneous introduction of multiple changes places more intensive demands on local and state staff compared to a phased implementation.

Figure 5.2 presents the three phases of the sequential phasing scenario, which begins with classroom PTELs.



**Figure 5.2 Sequential phasing scenario beginning with classroom PTELs**

**Scenario 3: Sequential Phasing Beginning with Modifications to End-of-Year Assessment**

Similar to Scenario 2, this scenario unfolds in three main phases (figure 5.3). To have the mini PT articulate with the common PTELs, the order of Phases 2 and 3 could be reversed.

**Phase 1**

* The end-of-year summative assessment is modified. This modification mirrors Phase 3 in Scenario 2, where the section containing discrete items is shortened and a mini PT is added to replace the existing CAST PTs\*.
* OTL indicators are identified in this initial phase.

**Phase 2**

* Similar to Phase 1 in Scenario 2, the curated set of classroom PTELs is developed and provided to teachers.

**Phase 3**

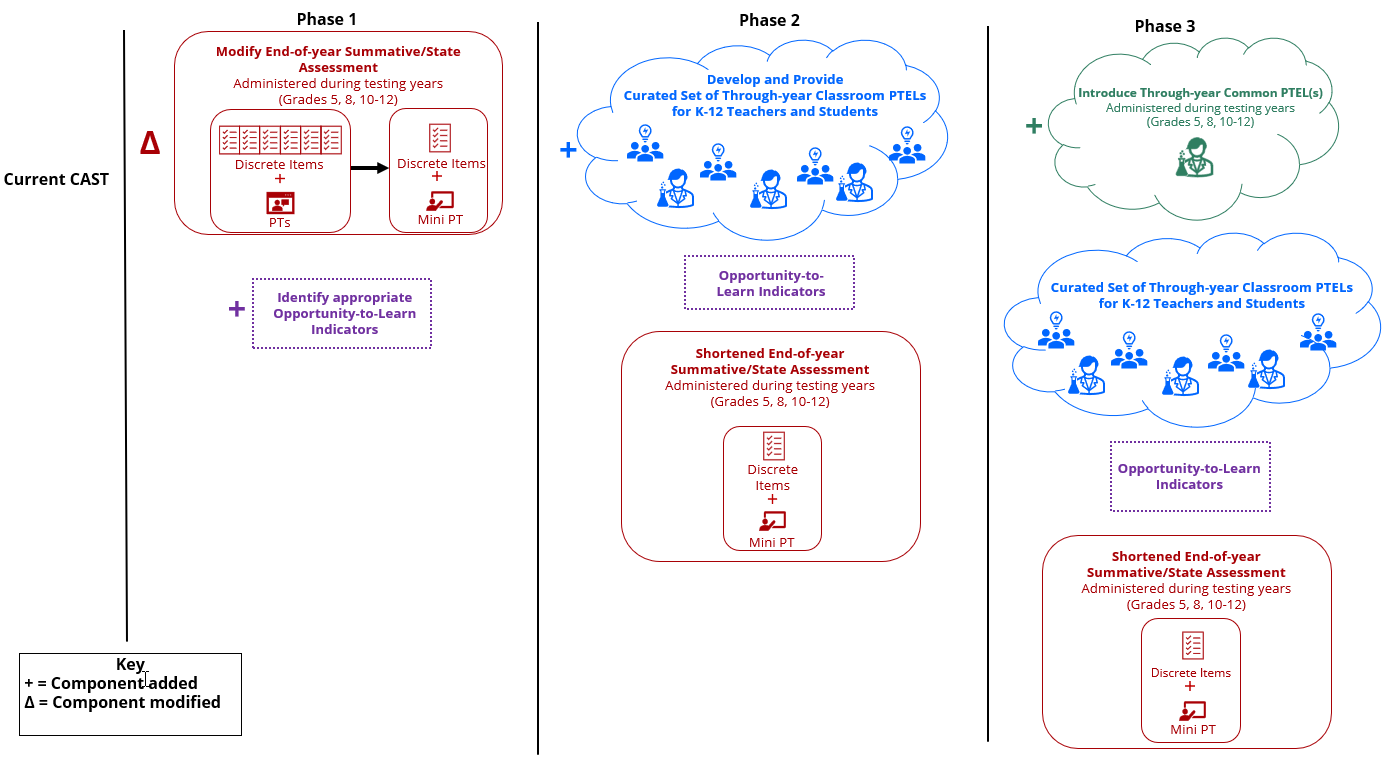
* Similar to Phase 2 in Scenario 2, the common through-year PTEL(s) will be developed and administered to students.

**Scenario 3 Analysis**

While Scenario 3 shares many of the advantages and disadvantages with Scenario 2 because of its sequential phasing, the order of the implementation differs and introduces several considerations.

This scenario allows for early-stage field testing of the modified end-of-year summative assessment (reduced discrete section and mini PT). The subsequent introduction of the curated set of classroom PTELs permits constructing this set to maximize synergy with the end-of-year mini PT. Introducing the components in this manner would also allow time to address potential issues in the administration and scoring of the common PTEL(s) in the final phase. One potential disadvantage of Scenario 3 compared to Scenario 2 is that the late introduction of the common PTELs will delay the availability of data needed to construct and evaluate a model for combining scores from them with the end-of-year summative assessment.

Figure 5.3 presents the three phases of the sequential phasing scenario, which begins with modifications to the end-of-year assessment.



**Figure 5.3 Sequential phasing scenario beginning with modifications to the end-of-year assessment**

**Scenario 4: Parallel Development**

This scenario emphasizes interconnections and alignment among various assessment components, particularly the two types of PTELs (figure 5.4). It involves two main phases:

**Phase 1**

* Classroom and common PTELs are designed together with a strong emphasis on coherence and interconnections such as similar tasks, shared data sources, and phenomena.
* OTL indicators are identified in a manner similar to the other scenarios.

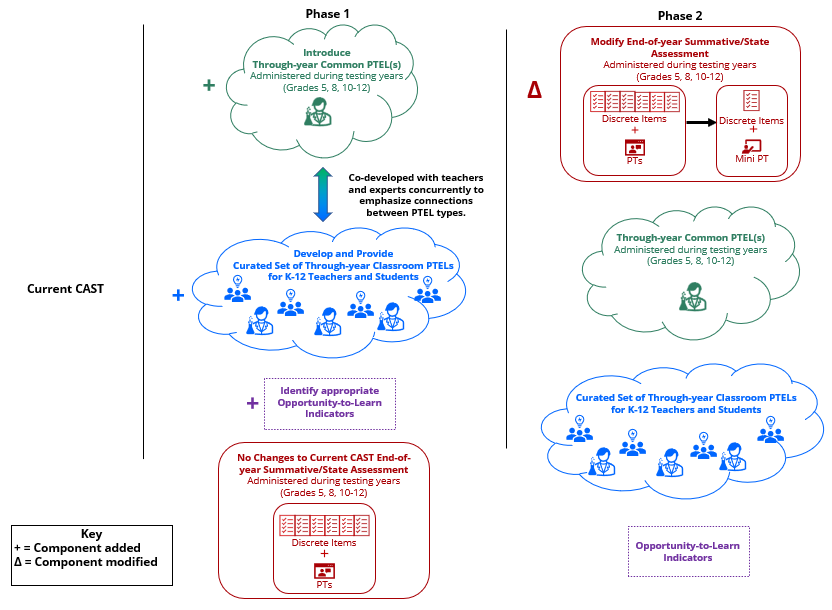
**Phase 2**

* Similar to Phase 3 in Scenario 2, the discrete items on the end-of-year summative assessment are reduced, and a mini PT is added to replace the existing CAST PTs.\*

**Scenario 4 Analysis**

This scenario shares the benefits of a more gradual approach as in Scenario 2 and Scenario 3 and the efficiency of a simultaneous implementation of some system components as in Scenario 1. Simultaneously developing and introducing the classroom and common PTELs allow for a closer consideration of the features that best support teaching and instruction. This scenario also provides opportunities to refine the procedures and methods for scoring the common PTELs before combining student performance with the score from the modified end-of-year assessment. Because the end-of-year assessment is modified in the second phase, the creation and evaluation of the summative scoring model would need to wait until after that implementation.

Figure 5.4 presents the two phases of the parallel development of the PTELs.



**Figure 5.4 Parallel development of the PTELs**

**Final Observation**

As briefly noted in the previous subsections, each of these scenarios has its advantages and challenges. The choice of which scenario to pursue depends, among other things, on the specific goals of implementation, resources available, ease of implementation, and the support available to help teachers adopt such changes.

**6.** **Future Directions and New Horizons**

There are several future directions that the CDE and SBE might monitor for potential incorporation in CAST innovations. Three of those directions are briefly described next.

**Diversity and Fairness**

In many countries, including the United States, population diversity has arguably reached unprecedented levels (Vertovec, 2023). Such diversity is most obvious not only in terms of race and ethnicity but also in native language, immigrant status, indicators of socioeconomic status, and other social characteristics. This shift has prompted a growing recognition of the need to reevaluate foundational assumptions about assessment to better align with principles of fairness, Universal Design for Learning, and equitable opportunities to learn. The overarching aim of such thinking is ensuring all students have the opportunity to demonstrate their knowledge and abilities fairly and equitably. Organizations, such as ETS, are actively exploring innovative approaches to address these issues.

**Artificial Intelligence**

The rapid public rise of large language models like ChatGPT has focused the interest of virtually every disciplinary community and industry on their application. Within the educational measurement community, AI technology has been in use for at least two decades, dating at least to the operational scoring of essays for the Graduate Management Admission Test in 1999. Most assessment companies are now experimenting with, if not already operationally employing, generative AI models in item creation, review, scoring, analysis, and reporting, typically in conjunction with human experts. The promise of these models is for increases in quality and speed, coupled with decreases in cost. One intriguing possibility is for the role these models could play in generating and scoring PTELs, including models that are more personalized to an individual.

**General Competencies Measured in Domain Contexts**

For some time, employers have been calling for workers with more developed durable skills—communication, collaboration, problem solving, critical thinking, and the like. A significant issue with respect to the measurement of those skills is that they can change as a function of domain. That is, although there may be common competencies across domains, the nature of critical thinking in science, for example (or collaboration, as a second example), is not identical to that in literary analysis or history. Because the CAST is a domain-based assessment grounded in the CA NGSS, and because the CA NGSS addresses some of these skills, a future direction might be to begin incorporating such competencies into the assessment system. The classroom PTELs and common PTELs would be the places to begin that exploration through, for example, tasks requiring collaboration. Scoring might then include aspects of collaboration for individuals and, separately, for the group.

**7. References**

Achieve. (2018). *Science task prescreen.* <https://nextgenscience.org/sites/default/files/resource/files/Achieve%20Task%20PreScreener_Final_9.21.18.pdf>

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>

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

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/>

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

Educational Testing Service. (2016). *Proposed design for California’s Next Generation Science Standards general summative assessments.* Educational Testing Service.

Educational Testing Service. (2019). *Use of evidence-centered design in CAST item and test development* [Unpublished report]. Educational Testing Service.

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

Fine, C., & Furtak, E. (2020). The SAEBL checklist: Science classroom assessments that work for emergent bilingual learners. *Science Teacher, 87,* 38–48.

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>

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

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/>

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>

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/>

Vertovec, S. (2023). *Superdiversity: Migration and social complexity*. Routledge.

**8. Accessibility Information**

**Alternative Text for Figure 4.1**

In the flowchart for the theory of action, there are four vertical lanes with headings. From left to right, the headings indicate that the first lane represents System Elements, the second represents Action Mechanisms, the third represents Intermediate Outcomes, and the fourth represents Ultimate Outcomes. Within each of these vertical lanes, elements of the theory of action are represented by shapes. A legend is presented at the bottom of the flowchart to identify the representation of the shapes. Components are represented by hexagons; Teacher is represented by trapezoids; Teacher and Student are represented by rectangles; Student is represented by parallelograms; and Local/State Educators, Policymakers, and Community Members are represented by ovals. The relationship among these elements is discussed in detail in the paragraphs preceding the image. The elements are listed next within each of their respective lanes.

**System Elements**

* CA NGSS Standards
* Professional Support for Educators in Implementing the System

*PT/PTEL System Elements*

* Through-Year Classroom PTELs

*Summative System Elements*

* Opportunity to Learn in Science Indicators
* End-of-Year Assessment
* Discrete

*PT/PTEL and Summative System Elements*

* Through-Year Common PTELs
* End-of-Year Assessment
* Mini PT

**Action Mechanisms**

* Teachers Emphasize Responsive Instruction around CA NGSS Science and Engineering Practices
* Teachers Incorporate Professional Support Experience into Practice
* Teachers and Students Use Timely Assessment Information to Make Inferences to Adjust Instruction and Learning
* All Students Engage in Authentic and Rigorous Science Learning and Assessment
* Local/State Educators, Policymakers, and Community Members Review the Results

**Intermediate Outcomes**

* Improved Teacher Understanding of CA NGSS
* Increased Responsive, Pedagogical Knowledge and Assessment Literacy among Teachers
* Instruction more focused on scientific reasoning and practices
* Better use of formative assessment
* Increased Student Sensemaking of Science Phenomena
* Better Use of Summative Results by Local/State Educators and Policymakers to Help Identify LEAs, Schools, and Classes in Need and Provide Effective Support
* More Informed Contributions by Community Members to Decisions Related to Providing Education Resources

**Ultimate Outcomes**

* Enhanced Teacher Practices
* Improved Student Learning and Motivation in Science
* More Effective Decisions by Educators and Policymakers
* Increased Community and Public Support for Providing Equitable Learning Opportunities

*(Return to figure 4.1.)*

**9. Appendix****: Development of the Proposed Conceptual Approach**

The conceptual approach described in [Section 3](#_A_New_Conceptual) emerged through a process of considering various approaches for integrating PTELs into an assessment system. In this appendix, ETS describes the four general approaches that were considered and the criteria that were used to identify the proposed conceptual approach as a reasonable option.

Table 1 provides a visual summary of the different assessment system components of the four approaches under consideration. All four conceptual approaches aim to integrate at least one of the two types of through-year PTELs (classroom or common) into the state summative assessment system.

Note the following for table 1:

* X = feature present.
* X:LE = feature present but would decrease or have less emphasis.
* The shading of Approach 3 denotes it as being the proposed approach described in this report.

**Table 1. Summary of the Components of Each Conceptual Approach**

| **Conceptual Approach** | **Component: Classroom PTEL for Instruction** | **Component: Through-Year Common PTEL** | **Component: End-of-Year Mini PT** | **Component: End-of-Year Discrete Items** | **Component: End-of-Year CAST PT\*** |
| --- | --- | --- | --- | --- | --- |
| **Approach 1:** Classroom PTELs for Instruction and End-of-Year Test for Summative Assessment | **X** | **-** | **-** | **X** | **X** |
| **Approach 2:** Classroom PTELs for Instruction and Through-Year PTELs plus End-of-Year Test for Summative Assessment | **X** | **X** | **X** | **X** | **-** |
| **Approach 3:** Classroom PTELs for Instruction and Through-Year PTELs plus End-of-Year PT and Reduced Discrete Items for Summative Assessment | **X** | **X** | **X** | **X:LE** | **-** |
| **Approach 4:** Classroom PTELs for Instruction and Through-Year PTELs for Summative Assessment | **X** | **X** | **-** | **-** | **-** |

**Conceptual Approach 1**

This approach maintains the existing CAST end-of-year summative assessment while introducing a curated set of classroom PTELs for instructional support. Teachers can choose PTELs relevant to their instructional units, and the end-of-year assessment includes both discrete items and the existing PTs. The classroom PTELs serve as a continuous learning resource and do not require formal scoring. Teachers receive professional development on PTEL selection and implementation and use student data from PTELs for instruction.

**Conceptual Approach 2**

This approach includes a curated set of classroom PTELs but adds common through-year PTELs that contribute to the summative accountability score. Along with the professional development to select and implement classroom PTELs, teachers participate in scoring the common PTELs, fostering shared understanding of student competence. Scoring processes are standardized for consistency.

**Conceptual Approach 3**

This approach is the proposed approach described in [Section 3](#_A_New_Conceptual) of this report.

**Conceptual Approach 4**

In this approach, through-year common PTELs entirely replace the end-of-year summative assessment and become the primary source for accountability. The scoring procedures may be similar to those used in Approach 2. This approach also includes a curated set of classroom PTELs.

**Considerations for the Assessment System to Support the Goals of Innovation**

To evaluate these four conceptual approaches, several criteria were considered. These criteria were identified on the basis of the current CAST program purposes and CAST innovation goals (i.e., to include the development of PTELs into the CAST and to shorten the summative testing time). The definition for each criterion is presented next:

1. **Local flexibility** refers to whether the PTEL innovation can be locally driven in its design, use and adaptation, and implementation.
2. **Comparability** **of evidence** sufficient to purpose and context of use refers to the degree to which each child has the best possible chance of demonstrating what the child can do, with the intent of optimizing construct-relevant variance for each individual.
3. **Test security** sufficient to the purpose and context for PTEL use, referring, for example, to the submitted work being that of individual students except for tasks allowing for collaborative work or access to external resources.
4. **Reduced end-of-year summative testing time** specifically refers to the impact on the length of the end-of-year summative assessment.
5. **Tractability of psychometric challenges** refers to the degree to which complexity in the measurement models relevant to the PTELs and their contribution to the summative scores can be managed.
6. **Sufficiency of standards coverage** that appropriately balances breadth and depth.
7. **Positive impact on teaching and learning** refers to the extent that PTELs accomplish the following:
8. Encourage the development of teachers’ competency in 3D science teaching and assessment
9. Are standards-aligned to ensure a strong connection between classroom and through-year or end-of-year summative PTELs to best foster students’ development of competencies
10. Are used because they are perceived as providing value that compensates adequately for any additional work required for teachers
11. Render information about student performance and progress that teachers can use for instructional planning and feedback
12. Include and engage all students in deeper learning
13. Support learning and assessment opportunities that are responsive to students’ backgrounds, experiences, identities, and prior knowledge
14. Allow for different modes of expression and problem representation so all students can demonstrate their performance
15. **Impact on the current development plan** refers to the extent of changes or additional new development work to be incorporated into the current assessment development plan.

When evaluated against the criteria listed, all four approaches provided local flexibility, standardization, and content coverage in summative assessment. Among the four conceptual approaches, **Conceptual Approach 3 appeared to be best.** It ***reduced total end-of-year summative testing time*** by shortening the discrete section and enhanced ***positive impact on teaching and learning*** by replacing the current PTs\* with mini PTs more likely to encourage teaching and learning of the CA NGSS scientific practices. Moreover, with the addition of the through-year classroom and common PTELs and the end-of-year mini PT, Approach 3 provided greater ***sufficiency of standards coverage*** by adding considerable depth. Additionally, Approach 3 appeared to fare well on ***local flexibility,*** giving schools and LEAs the freedom to decide when to administer the through-year common PTEL(s) based on their annual plan or instructional sequences. It did well on ***comparability of evidence*** in that it maintained an end-of-year assessment coupled with the standardized common PTEL(s).

While Conceptual Approach 3 offers several advantages, it may be lower than some other approaches on the ***tractability of psychometric challenges*** due to the need for innovative psychometric models that allow for the combination of students’ performance on the common PTEL(s) with end-of-year summative scores. Furthermore, Approach 3 has a somewhat greater ***impact on the current development******plan*** than other approaches. These limitations, however, are arguably outweighed by their advantages.