California Department of Education

Executive Office

SBE-002 (REV. 11/2017)

memo-imb-adad-dec19item01

# **MEMORANDUM**

DATE: December 9, 2019

TO: MEMBERS, State Board of Education

FROM: TONY THURMOND, State Superintendent of Public Instruction

SUBJECT: Update on the California Science Test.

## Summary of Key Issues

This Information Memorandum provides updates on the results from studies and analyses conducted by testing contractor Educational Testing Service (ETS). These updates include: the multistage adaptive test (MST), the screener, the use of matrix sampling, and the time needed for students to complete the California Science Test (CAST). For these studies and analyses, ETS used the results from the CAST 2018–19 first operational administration. In January 2020, using the results from ETS’s report *Informing the California Science Test Blueprint Improvements: Results from the Psychometric Studies*, found in Attachment 1, the California Department of Education (CDE) will present the recommendation to revise the CAST blueprint to the California State Board of Education (SBE) for approval.

## California Science Test Development Background

As California worked with ETS to develop the CAST, it gained recognition as a national leader for embracing the challenging Next Generation Science Standards. In accordance with California *Education Code* Section 60640(b)(2)(B), the development of the CAST design was based on consultation with national science experts, including Jim Pellegrino, Helen Quinn, Kathleen Scalise, and Jonathan Osbourne, and extensive feedback from California educators and other stakeholders. Using this information, CDE leadership and ETS’s assessment design team created an innovative test design. The SBE approved the CAST high level test design (HLTD) to measure the full range of the California Next Generation Science Standards (CA NGSS) in March 2016.

Table 1 presents a summary of the SBE-approved CAST HLTD that displays the three segments: Segment A, Segment B, and Segment C. Briefly, Segment A consists of stand-alone items;[[1]](#footnote-1) Segment B consists of two performance tasks;[[2]](#footnote-2) and Segment C consists of either a block of stand-alone items or another performance task that could be used for matrix sampling[[3]](#footnote-3) and field testing. For a comprehensive overview of the CAST design, refer to the August 2018 Memorandum, which is located at <https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-adad-aug18item01.docx>.

Table 1. CAST Design Summary

\*Also includes the foundational concepts introduced in kindergarten through grade two.

| Characteristics | Segment A | Segment B | Segment C |
| --- | --- | --- | --- |
| Reporting level | Contributes to student and group reports | Contributes to student and group reports | Contributes to group reports |
| Scope and depth of measurement | Wide breadth—measures a broad sample of performance expectations (PEs) | Deep measurement of targeted sample of a few PEs provided in item sets | Broad and deep—full range of measurement of PEs for each grade span |
| Number and type of items | 32–34 stand-alone items that include selected-response, technology enhanced, machine-scorable items | Two performance tasks (with 4–6 item sets) | 12 or 14 stand-alone items that include selected-response, technology enhanced items (like Segment A) or one performance task (like Segment B, with 4–7 item sets) |
| Grade by PEs measured  | Grade 5 assesses PEs\* from grades 3–5Grade 8 assesses PEs from grades 6–8High school assesses PEs from grades 9–12  | Grade 5 assesses PEs\* from grades 3–5Grade 8 assesses PEs from grades 6–8High school assesses PEs from grades 9–12 | Grade 5 assesses PEs\* from grades 3–5Grade 8 assesses PEs from grades 6–8High school assesses PEs from grades 9–12 |

The CDE and ETS were committed to providing periodic updates on the test development, which included refinements of the approved design plan in recognition of the innovative approach and complexities of the standards. To provide evidence of the utility of the HLTD, ETS conducted studies investigating the practicality of the MST[[4]](#footnote-4) and screener.[[5]](#footnote-5) In addition, ETS performed analyses of the time needed for students to complete the CAST.

In November 2017, the SBE approved the CAST blueprint. In accordance with the approved blueprint, the CAST will measure all CA NGSS PEs for grades three through twelve over a three-year cycle.

In its August 2019 Memorandum, the CDE updated the SBE that the 2019−20 CAST form will continue to use the 2018−19 format as the test forms needed to be finalized before the testing window closed for the 2018–19 administration. That Memorandum is located at <https://www.cde.ca.gov/be/pn/im/documents/memo-pptb-adad-aug19item01.docx>.

## California Science Test—Results from Studies

ETS and the CDE continue to collaborate with educators, national science experts, various stakeholders, and the California Assessment of Student Performance and Progress (CAASPP) Technical Advisory Group (TAG) in developing and administering the CAST. Using the CAST results from the first operational year, ETS conducted studies and analyses regarding the MST, screener, and how much time students are taking to complete the CAST.

### Multistage Adaptive Test Study

The CAST HLTD includes the implementation of an MST. The following sections detail the background and findings from the multistage adaptive test study conducted by ETS.

#### Multistage Adaptive Test Study Background

An MST features two or more components: the first stage, which is administered to all students and covers the full range of possible performance; and a second stage (or more), which is selected on the basis of student performance on the prior stage. For example, a student who performs extremely well on the first stage of the assessment will receive test items in the next stage of the test that are at a higher level of difficulty. This type of assessment can improve measurement, at the tails (i.e., the extreme ends of the ability distribution), by better targeting items to student achievement and requires a sufficiently large item pool to support the varying components.

#### Multistage Adaptive Test Study Findings

The MST study evaluated the feasibility of structuring the assessment in a stage-adaptive format for Segment A and whether implementing the MST would offer improved measurement precision over that of a fixed-form test. However, results of the study revealed that, for students at the extreme ends of the ability distribution (i.e., students who performed very well or poorly on the CAST), the MST form does not provide significantly more information about these students’ performance. The findings indicate that, at this time, the number of CAST items at the higher difficulty range (i.e., higher difficulty range means items are easier) is insufficient to support the building of the second-stage block of items for the MST.

### Screener Study

The CAST HLTD also includes a screener. The following sections detail the background and findings from the screener study conducted by ETS.

#### Screener Study Background

The screener was theorized to be used as a student transitions from Segment A (stand-alone items) to Segment B (two performance tasks). The impetus for the screener was to increase the precision of measurement by assessing students in the science domains they exhibited relative strengths, thereby collecting in-depth information about a student’s knowledge, skills, and abilities in the assessed standard(s).

The student’s performance in Segment A may determine the science domains (i.e., Earth and Space Sciences, Life Sciences, and Physical Sciences) the student will be presented in Segment B. For example, if a student performs significantly lower in a particular science domain (e.g., Physical Sciences) compared to that student’s performance in the other science domains, the screener eliminates that science domain performance task when the student transitions to Segment B; that student would then be presented one Life Sciences and one Earth and Space Sciences performance task instead. If a student were to perform relatively equally in all three science domains in Segment A, the assignment of the two performance tasks presented in Segment B would be randomly assigned.

#### Screener Study Findings

The screener study explored the potential utility of implementing a screener. The results indicated that the precision of measurement of students’ performance was not improved by the use of the screener algorithm.

### Matrix Sampling

The 2017 SBE-approved CAST blueprint states that Segment C consists of either a block of stand-alone items or another performance task, both of which could be used for matrix sampling and field testing. If Segment C were used solely for field testing, until the item pool is built to a sufficient volume, a change to the CAST blueprint would be required.

### CAST Administration Time

ETS analyzed how long students took to complete the 2018–19 CAST. The total expected completion time for the CAST is approximately two hours: 60 minutes for Segment A, 40 minutes for Segment B, and 20 minutes for Segment C. The results of the analysis indicated that a high percentage of students in grades five and eight did not complete the test within two hours. The data on student testing times suggests that an adjustment to the blueprint, or an adjustment to the testing time expectation, needs to be made.

### Group-Level Reporting

The CDE will post group-level reports on the public reporting website in January or February 2020. These reports will be by grade and by student groups at the state, county, district, and school levels. In addition, the CDE and ETS are continuing to collaborate and explore, with national science experts, options to expand group-level reporting.[[6]](#footnote-6) For example, targeted reports may provide additional information to educators about their students’ relative strengths and areas of weakness to inform teaching and learning. The CDE will provide information to the SBE at the completion of these analyses as well as any additional recommendations that may be required.

### Next Steps

In January 2020, the CDE will provide recommendations and propose actions to the SBE on the CAST design, including the blueprint. The CDE will keep the SBE informed of on-going analyses, studies, and developments for the CAST.

## Attachment(s)

* Attachment 1: Informing the California Science Test (CAST) Blueprint Improvements: Results from the Psychometric Studies (182 Pages)
1. A stand-alone item (i.e., discrete item) assesses a student’s knowledge, skill, and abilities on a standard or part of the standard. Stand-alone items consist of multiple-choice, constructed response, technology enhanced, or composite items, where composite items are a combination of two item types; for example, a multiple-choice part and technology enhanced part to make one item. [↑](#footnote-ref-1)
2. A performance task presents a scenario or situation (i.e., stimulus) that requires students to apply their knowledge, skills, and abilities to respond to a series of questions that align with two or more related standards. [↑](#footnote-ref-2)
3. Matrix sampling in assessments is a methodology used to enhance the efficiency of measurement. In conventional assessments, every student completes every item on the assessment. However, in matrix sampling, each student completes only a sample of the items, thereby allowing the assessment to assess a wider range of items while reducing testing time. [↑](#footnote-ref-3)
4. An MST is a compromise between a traditional fixed-form test (e.g., all grade five students are given the same test questions or items) and a computer adaptive test (CAT). Rather than administering a single fixed-form test or adapting the test to individual students item by item, as in a CAT, an MST adapts to students in stages. [↑](#footnote-ref-4)
5. The screener is used as each student transitions from Segment A to Segment B on the CAST. The student’s performance during Segment A may determine the science domains (i.e., Physical Sciences, Life Sciences, and Earth and Space Sciences) the student will be presented in Segment B. [↑](#footnote-ref-5)
6. Group-level reporting provides feedback needed for program evaluation and promotes curriculum improvement while, at the same time, ensuring that each student is measured fairly and comparably. [↑](#footnote-ref-6)