Embedded Field Test of Performance Tasks Design and Talking Points

Background

Large-scale assessment systems, like Smarter Balanced, draw test items from an item bank. To maintain a current, high-quality item bank, new items must be field-tested. The field-tested items are embedded in operational tests. Students do not know the difference between an operational item and a field-test item when they are taking the test. Student responses to the field-test items help Smarter Balanced further refine and eventually deploy the field-tested items as operational items. All items on operational Smarter Balanced ELA/literacy and mathematics tests, including performance tasks (PT) and computer adaptive test (CAT) items, are field-tested prior to appearing on the test as operational items.

Performance tasks are critical elements of the Smarter Balanced summative and interim assessments. The Smarter Balanced Technical Advisory Committee described that the performance tasks serve important purposes. They assess elements of content that are necessary to support making valid inferences regarding student performance and they signal high-quality instruction.

Smarter Balanced has developed new performance tasks that incorporate improvements based on the lessons we learned from earlier development efforts. These performance tasks were designed at the outset with accessibility resources that students need to demonstrate their knowledge and skills. In addition, these new performance tasks are not associated with any classroom activities.

Consortium members investigated several possible methods of field-testing new performance tasks. Based on numerous discussions, members determined that an embedded field test of new performance tasks best addresses the goals of minimizing additional student testing time and maximizing the quality of data that are derived from the new performance tasks.

The Challenge of Field-Testing Performance Tasks

In spring 2017, Smarter Balanced will be field-testing ELA/literacy and mathematics CAT items across member states/territory, which together test approximately six million students. In addition, each embedded field-test performance task (EFT PT) is pre-assigned to about 2,000 students consortium-wide. This means that the number of students in each member location receiving an EFT PT will be relatively small.

Administering field-test performance tasks, in addition to operational performance tasks, would have increased testing time for students. To minimize this issue, the Consortium agreed to randomly select a small percentage of students to receive a field-test performance task instead of an operational performance task. These students will be presented with a small number of additional items on the computer adaptive portion of the test. The “one performance task per student" design will ensure that no student will be unduly burdened with extra testing time.

What Does This Mean for Students?

Field-test items do not contribute to a student's total score. Students receiving a field-test performance task will only receive scores for operational items. It is possible to generate a reliable and accurate overall scale score, even when a performance task does not contribute to that score. The students who are assigned a field-test performance task will, however, have the same overall testing experience as students who are assigned an operational performance task.

Roles and Responsibilities

Smarter Balanced members and staff share a vested interest in ensuring the successful field-testing of performance tasks.

Members are responsible for:

- Administering the tasks in a manner consistent with the Smarter Balanced Online Summative Assessment Test Administration Manual, and
- Providing the student responses associated with the EFT PTs as well as the student responses to all other items to the Smarter Balanced scoring contractor.

Smarter Balanced is responsible for:

- Developing the PTs and the associated CAT test packages,
- Contracting with a vendor for scoring, range finding, conducting data reviews, and calibrating the EFT CAT and PT items,
- Producing anchor papers to support 2018 operational hand scoring, and
- Alerting members of sensitive responses to field-test performance task items per the guidance provided in the *Smarter Balanced Online Summative Assessment Test Administration Manual.*

Q&A

1. If a student can receive a reliable, accurate score from only the computer adaptive test, why would we administer a performance task as part of the test?

Without a performance task, we can assert that a student's score reliably shows how much of the subject matter the student has learned. It does not invalidate the score. But a score from an operational assessment that includes a performance task provides unique information about what a student knows and is able to do.

There are also instructional implications if performance tasks are not included on the assessment. The Smarter Balanced performance tasks are developed using guidelines that mirror good instructional practices. Each performance task must:

- Integrate knowledge and skills across multiple content standards or strands within a content area.
- Measure capacities such as depth of understanding, research skills, complex analysis and identification/provision of relevant evidence to support an argument.
- Require student-initiated planning, management of information and ideas, and interaction with other materials.

- Require the production of extended responses, such as essays.
- Reflect a real-world task and/or scenario-based problem.
- Allow for multiple approaches.
- Represent content that is relevant and meaningful to students.
- Allow for the demonstration of important knowledge and skills, including those that address 21st-century skills such as critically analyzing and synthesizing information presented in a variety of formats, media, etc. (Source: Smarter Balanced Performance Task Specifications)

Including performance tasks on the assessment supports the use of these practices during classroom instruction.

2. Which mathematics and ELA/literacy grades are involved?

The 2016-17 summative test packages include embedded field-test performance tasks in grades 3–8 and high school for ELA/literacy and mathematics.

3. How many performance tasks will be included in the spring 2017 embedded field test for each grade and content area?

There will be five (5) ELA/literacy performance tasks and four (4) mathematics performance tasks in each grade.

4. How many students will be assigned a field-test performance task in each grade and content area?

The table below presents the percentages of students who will be assigned a field-test performance task for each grade and content area. These percentages will be applied to each member's tested student population to calculate the number of students who will be assigned a field-test performance task.

Content Area	Grades	Sampling Rate
Mathematics	3 – 8	1.2%
Mathematics	High School	1.4%
ELA/literacy	3-8	1.5%
ELA/literacy	High School	1.7%

Smarter Balanced needs about 2000 student responses consortium-wide for each field-test performance task.

5. Why is the percentage of students higher for ELA/literacy than mathematics?

Because we will be field-testing five (5) performance tasks in ELA/literacy and four (4) performance tasks in mathematics in each grade, the sampling rate is slightly higher in ELA/literacy than in mathematics.

6. Why is the percentage of students higher for high school than for grades 3-8?

The sampling rate is slightly higher in high school as compared to grades 3–8 because fewer students take the Smarter Balanced assessments in high school as compared to grades 3–8.

7. How many additional items will be included on the computer adaptive test (CAT) portion for students who receive a field-test performance task?

- For mathematics, two additional items per grade level will be included.
- For ELA/Literacy, three additional items will be included.

8. How will these additional items appear in students' tests?

The additional items will be embedded in the CAT portion for those students who are assigned a field-test performance task.

9. How much time will students need to answer the additional items?

On average, it will take approximately 10–15 additional minutes for students to answer these items.

10. Why does the design include additional items in the CAT for students who are assigned a field-test performance task?

Some discussion was given to the idea of including the additional items at the end of the field-test performance task, but there were two concerns: (1) it would be difficult to communicate to teachers regarding the inclusion of stand-alone items on the field-test performance tasks, and (2) the presence of additional stand-alone items on a performance task would signal to students that they were assigned a field-test performance task.

11. How will claim scores be determined for students who are assigned a field-test performance task?

The additional items included in the CAT portion together with the other CAT items will be used to support reporting on the claims for the small percentage of students across the Consortium who are assigned a field-test performance task.

12. Will Smarter Balanced provide writing trait scores for students who receive an ELA/literacy EFT PT?

Smarter Balanced will provide members with writing trait scores for ELA/literacy full write items after they are subsequently scored and calibrated.

- Members may opt to score the ELA/literacy PTs in parallel to expedite the reporting of writing trait scores.