Mathematics Performance Task Specifications

Revised February 2017

TABLE OF CONTENTS

DESIGN SPECIFICATIONS

A Performance Task (PT) is an item type designed to provide students with an opportunity to demonstrate their ability to apply their knowledge and higher-order thinking skills to explore and analyze a complex, real-world scenario. A mathematics performance task elicits evidence of students' ability to "bring it all together" to develop a solution plan to the central challenge of the task.

The first section of this document defines A) key features, B) gatekeeper criteria, and C) criteria descriptors to guide the development and review of performance tasks for the Smarter Balanced Assessment Consortium. Task types (equivalent to genres in ELA) in mathematics include *Plan and Design*, *Analysis and Theory* and *Evaluate and Recommend* and hybrids of these types. Technical considerations for each task type are found in sections D–K of this document. The design specifications in this document are for performance tasks in mathematics and work in tandem with other *Consortium*-approved specifications, including mathematics content specifications, grade-level areas of emphases, universal design, general accessibility and bias guidelines, metadata requirements, and style guidelines for mathematics text, equations, and artwork.

A. Key Features of Mathematics Performance Tasks

(adapted from *Smarter Balanced Performance Task Specifications* document)

Performance tasks should:

- integrate knowledge and skills across multiple Claims and Targets–a key component of college and career readiness.
- measure capacities such as depth of understanding, research skills, and/or complex analysis with relevant evidence.
- require student-initiated planning, management of information/data and ideas, and/or interaction with other materials.
- reflect a grade level, developmentally appropriate real-world problem. Tasks elements (data sets, values provided, etc.) are not restricted to those actualized in the real world, but these elements should be realistic.
- allow for multiple approaches.
- represent content that is relevant and meaningful to students.
- allow for demonstration of twenty-first-century skills, such as critically analyzing and synthesizing information presented in a variety of formats, media, etc.
- require scoring that focuses on the essence of the Claim(s) and Targets for which the task was written.
- be feasible for the school/classroom environment.

B. Gatekeeper Criteria for Performance Assessments

A set of cross-cutting criteria have been developed to guide the development and review of Performance Tasks. The term *gatekeeper* indicates these as essential components of a quality PT. Performance Tasks that do not meet these criteria would not be accepted into the PT item pool. The gatekeeper criteria are listed below.

Aligned with Claims and Standard8s	PTs should go to the heart of the key Smarter Balanced Claims and Common Core State Standards for Mathematics. In particular, they should elicit evidence of Claims 2, 3, and 4: • Students can solve a range of complex, well-posed problems in pure[1] and applied mathematics, making productive use of knowledge and problem solving strategies. (Claim 2) • Students can clearly and precisely construct viable arguments to support their own reasoning and to critique the reasoning of others. (Claim 3) • Students can analyze complex, real-world scenarios and can construct and use mathematical models to interpret and solve problems. (Claim 4) In addition, they should elicit evidence of student engagement in the Common Core Mathematical Practices.
Developmentally Appropriate	PT topics, tasks, and scoring should be appropriate for the age and developmental experience base of the students.
Engaging	Topics should be authentic and realistic, engaging students in solving a problem or making a decision they would find relevant.
Accessible	Topics and tasks should minimize sources of bias, allow for multiple pathways, and provide appropriate scaffolds or supports while keeping in mind that sources and response types need to allow access for students with different English language proficiency and students with disabilities.
Purposeful and Coherent	Tasks should have an authentic purpose, and all task components should be connected to achieving that goal.

[1] Note: PTs in particular aim to assess problem solving in applied mathematics.

C. Criteria Descriptors

Performance tasks should go to the heart of the Smarter Balanced Claims and Common Core State Standards for Mathematics (CCSSM) and should follow the criteria described below as they pertain to the choice of topic, task, and scoring.

Topic/Setting/Scenario

Developmentally Appropriate (age, grade, complexity, and content).
Driving Question: Is the task matched to the appropriate level for students?

- The topic should be developmentally appropriate. That is, it should be based on direct experiences the students will likely have had at this age level. This also requires consideration of what students from different cultural and economic backgrounds may have experienced.
- The topic should also be framed in a way that is appropriate to the cognitive, social, and emotional levels of the student (e.g., more concrete in earlier years; not emotionally disturbing for the age level of the students).

Engaging (relevant, of interest, builds on curiosity)
Driving Question: Would students want to do the task?

- The topic should be engaging to students. That is, the topic should be relevant to concerns that are typical of students at this age, deal with activities they would likely engage in or issues that would be of interest to many students.
- The topic should build on the natural curiosity of students.

Accessible (minimize bias, linguistic complexity, open-ended)
Driving Question: Can students access the task?

- The topic should be accessible to students with a wide range of experiences and minimize bias that could disadvantage a particular socio-economic, cultural-, or gender- group, or one living in a particular geographical location.
- Allowing choices with respect to how a student would approach a topic/task is one way to increase accessibility.
- Given that the context of PTs is rich and complex, any relevant hidden rules, underlying assumptions, background knowledge, or construct-relevant vocabulary related to the context should be made explicit to the student within the task.

Purposeful and Coherent
Driving Question: Do students have a coherent goal or purpose?

- The topic of the PT should offer a clear purpose and an authentic challenge in which the student must accomplish a meaningful goal.
- The scenarios for the performance task should provide a realistic, engaging, and authentic purpose for responding to the prompts.

Task: Student Role/ Stimulus/Prompts

Developmentally Appropriate (age, grade, complexity, and content).
Driving Question: Is the task matched to the appropriate level for students?

- The task should be developmentally appropriate to a range of students and use content drawn from earlier grade levels (up to the current grade level) so that students can focus on the complex application of higher-order skills (problem-solving, communicating reasoning, modeling, and analyzing data).
- The student's role in the scenario should be appropriate for the target grade level, age, and content.
- The prompts should be developmentally appropriate in complexity (i.e., reflect how much articulation of thinking a student can do). The expectation of explication should be developmentally appropriate.

Engaging (relevant, of interest, builds on curiosity)
Driving Question: Would students want to do the task?

- The student should have a leading role in the task to develop a plan, model, design, critique, and/or recommend.
- Within the given constraints, students should have a set of choices and decisions to make, and they should have an opportunity to follow through on the implications of their choices.
- The task should build on the natural curiosity of students at their age span or engage them in a question that is likely to evoke genuine interest or provide a familiar context in which they can relate. Tasks should not be trivial, uninteresting to a student at this age, or obscure.

Accessible (minimize bias, linguistic complexity, open-ended)
Driving Question: Can students access the task?

- The task should allow for a range of responses in depth, complexity, and mathematical sophistication.
- As students weigh evidence and make choices, the task should allow for a range of acceptable answers that are appropriate to the choices they make.
- The prompts should allow for multiple ways for students to solve the problem and to display their reasoning (i.e., textual explanations, graphs, equations, tables).
- The collection of sources should be sufficiently varied in format (including texts with a reading level at or below grade level, tables, data displays, and other visual sources) to allow for different entry points into the mathematics.
- To allow for multiple ways of representing student thinking, tasks can allow for choices in the use of technology-enhanced tools (graphing a line, manipulating an object, building a picture graph). At the same time, such tools should not be required unnecessarily, as they raise accessibility concerns for students unfamiliar with technology (these will typically be students in under-resourced schools) or students for whom ease of use is an issue (e.g., visually impaired, motor skills).
- The task should be worded in clear, simple, and unambiguous ways (avoiding the use of unfamiliar or unessential vocabulary and overly complex sentence structures. [2]
    - However, the task should include sufficiently clear and vivid language to convey the richness of the context or complexity of the scenario. Language can be made more accessible by ready access to visual representations and definitions of words or phrases (e.g., by allowing students to click on words for a definition).
- The task should clearly request the information on which scoring will depend.
- For construct irrelevant terms that need defining, tasks should include a pop-up glossary.

[2] Refer to Table 1. Language Complexity Rubric in WCER Working Paper No. 2013-5 available at http://www.wcer.wisc.edu/publications/workingPapers/Working_Paper_No_2013_05.pdf for guidance on text density, language form and structure, and vocabulary.

English language learner considerations:

- Provide directions that maximize clarity and minimize the potential for confusion.
- Use vocabulary that is widely accessible to all students, and avoid homonyms and unfamiliar vocabulary that is not directly related to what is being measured.
- Keep sentence structures as simple as possible while expressing the intended meaning.
- Do not use cultural references that are not equally familiar to all students or idiomatic expressions.
- Avoid sentence structures that may be confusing or difficult to follow, such as the use of passive voice, sentences with multiple clauses, or double negatives.
- Minimize the use of low-frequency, long, or morphologically complex words and long sentences.

Purposeful and Coherent

Driving Question: Do students have a coherent goal or purpose?

- The task should communicate to the student the overall purpose of the task and what they will be expected to accomplish.
- The prompts should be clearly related to accomplishing the overall purpose of the task and remain within the context of the given scenario.
- The prompts should follow a logical progression, building toward the culminating prompt (final design, written evaluation, or recommendation).
- The culmination of the task should represent a feasible solution to the overall problem in the scenario.

Scoring

Developmentally Appropriate (age, grade, complexity, and content).

Driving Question: Is the task matched at the appropriate level for students?

- Scoring guides should use criteria and expectations geared to the developmental level of the students.
- For example, elementary-grade level students' explanations of their reasoning should not be expected to be highly formalized and fully elaborated. (Students are still acquiring the language of mathematics.)

Accessible (minimize bias, linguistic complexity, open-ended)

Driving Question: Can students access the task?

- Scoring guidelines should be based on information directly requested in the prompt. (Do not score on what you did not ask.)
- Scorers should be instructed to look for the mathematical meaning of students' explanations rather than the way (grammar, syntax) they express themselves in Standard English.
- Where permitted by Smarter Balanced guidelines, scoring guidelines should allow for the use of other languages in their responses.
- Scoring guides should provide a full range of sample responses to account for multiple solutions, solution strategies, and/or explanations of reasoning.
- Scoring guides should allow for the awarding of partial credit.

Purposeful and Coherent

Driving Question: Do students have a coherent goal or purpose? To reflect the fact that students

are completing a coherent task with multiple parts, the following scoring guidelines should apply:

- Students should not be penalized on multiple prompts for errors that occur in one answer and are reflected in later answers.
- In cases where a response to one item is used to respond to another item, students should be awarded credit for correct reasoning.

TASK DEVELOPMENT AND SCORING SPECIFICATIONS

Task writers should ensure that the mathematics of the task is correct, and should use precise mathematical language.

D. Task Type Considerations

Plan and Design:

- For plan-and-design tasks, design constraints and parameters should be clearly labeled and explained in the prompt.

Evaluate and Recommend:

- For evaluate-and-recommend tasks, data sets should be drawn from authentic data sets. If they are mocked up, they should conform to reasonable estimates.
- The audience and format of the recommendation should be clearly stated (e.g., "Write a letter to your school principal with your recommendation.")

Analysis and Theory:

- These tasks provide a natural opportunity to engage students in the refinement of their theories, as additional data can be provided which suggests another theory.

E. Blueprint

1. Stimuli Presentation

When presenting stimuli, the following guidelines apply:

- There should be a reference (using bold font as an indicator) that connects items to the specific stimulus resource required for a response. The name of the resource should be bold in both the stem and stimulus (e.g., Use Table 1 to answer this question).
- The number of resources (tables/graphics) within the stimulus of a PT should be limited for grades 3–5, as follows:
  - Grade 3: two or fewer
  - Grades 4 and 5: three or fewer

2. Sets of Items

Each Grades 3–7 Performance Task should consist of a set of four to six items and each Grades 8 and 11 Performance Task should consist of a set of three to five[3] items (total points not critical). Each item may be worth 0–4 points, with guidelines for awarding 0, 1, 2, 3, or 4 points, but no half points. Scoring guides should allow for partial credit.

[3] At grades 8 and 11, a hand scored proof or justification extended response item aligned to Claim 3 will be included in the examination.

3. Ramping and Decision Making

Grades 3–7 PTs should exhibit appropriate ramping across items, while higher grades will exhibit less ramping with each increasing grade. A ramped sequence of items should provide all students access to a task, while maintaining the challenge of the stated goal of the task for higher performers. A ramped sequence of items will begin with items that encourage entry into the task but are still directly related to the stated goal of the task. All PT items should focus on problem solving, reasoning, and modeling as opposed to procedural and computational skills.

While ramping is an appropriate part of building toward autonomous un-fragmented chains of reasoning in Grades 3–7, by Grades 8 and 11, tasks should utilize less ramping, be less closely guided, and require more autonomous decision-making.

The number of items in a PT should correspondingly decrease as the grade level increases, indicated by the aforementioned decrease from 4–6 items per PT in Grades 3–7 to 3–5 item per PT in Grades 8 and 11.

4. Independent vs. Interdependent Items

Guidelines for the first 1–2 items:

a. The first 1–2 items must be independent (i.e., not needed to score subsequent parts of the task), but still directly related to the stated goal of the task.
b. The first 1–2 items should provide entry into the task. Here, "entry" means having low to low-medium difficulty and encourage students to make sense of the stated goal of the task.

Guidelines for the remaining items:

a. The remaining items in all performance tasks may be hand-scorable.
b. The remaining items cannot depend on the independent item(s) mentioned above.
c. The remaining items should be cohesive and may be interdependent with each other.
d. Rubrics for the interdependent items should explicitly prevent students from being penalized multiple times if a mistake made on a preceding item is correctly carried through subsequent items.
e. At grades 3–7, a minimum of two and a maximum of four items should be hand scored.
f. At grade 8 and 11, a minimum of one and a maximum of four items should be hand scored.

5. Alignment to Claims

Performance Task items should contribute scores to each claim according to the following distribution. Again, there should be a total of 4–6 items in grades 3–7 PTs, and 3–5 items in grades 8 and 11 PTs.

Claim/Score Reporting Category	PT Items
1. Concepts and Procedures	0
2. Problem Solving	1–2
3. Communicating Reasoning	1–2
4. Modeling and Data Analysis	1–3
Total	3–6 items

CHANGE LOG

Date	Page Number	Description	Rationale
2/17/2017	8	Addition of "Analysis and Theory" section	Provides further explanation of possible task types
2/17/2017	8	Removal of Classroom Activity	No longer needed for PTs
2/17/2017	9	Addition of "Ramping and Decision Making" section - number of total items in the PT in grades 8 and 11 was reduced to 3-5 from 4-6	Explains greater need of scaffolding of tasks in lower grades compared to upper grades
2/17/2017	9	In the "Independent vs. Interdependence" section, guidelines for the first "two" items is changed to the first "1-2" items	Less of a need at higher grade level for "entry" level, independent items