

Smarter Balanced Assessment Consortium: Technology-Enhanced Items Guidelines

Developed by Measured Progress/ETS Collaborative
April 16, 2012



Introduction

Technology-enhanced items (TEI) are computer-delivered items that include specialized interactions for collecting response data. These include interactions and responses beyond traditional selected-response or constructed-response. A TEI template describes a single interaction, response data collected as a result of that interaction, and the logic applied to score the response data. TEI templates can be used across grade levels and content areas. An important TEI design principle is to adopt innovative technology designs in concert with measurement procedures for effective validation and use of TEI. As new technologies become available, they can be adopted when required for students to produce new forms of evidence.

This directory contains: a) an introductory description of key terms and definitions related to TEI; b) a resource to help item writers choose TEI templates based on the evidence students are required to produce to support a target; and c) the collection of 25 TEI templates developed for this project. Each template is described in a template specification form that contains a description of the interaction and the parameters and data that describe the interaction and the scoring of student responses.

The template specifications contain example items to illustrate the concepts related to the template (e.g., interaction parameters, response data, etc.). These sample items are **not** aligned to specific assessment targets and are only meant to illustrate how an item writer would use the template to specify item parameters. Some of these sample items include tolerances, designed to show how an item writer can use parameters to account for imprecision in the test-taking interface. The tolerance values in the sample items were arbitrarily defined and are not intended to be incorporated into actual items; appropriate tolerances should be determined through cognitive labs and/or field testing.

The template specifications include interaction instructions. These instructions provide detailed, step-by-step descriptions of how the student would interact with an item developed from the template. These instructions are intended to inform the training of students *prior* to the assessment administration. Instructions may need to be modified based on the delivery system. Further, the instructions should be modified based on results of cognitive labs designed to evaluate the level of detail required in the instructions. In addition to the interaction instructions, each template includes sample item directions that are appropriate for direct inclusion in the item text.

Some specifications describe accessibility challenges that might arise in developing items from the template. Future work will need to address these challenges.

A subset of templates was selected for the initial operational stages. For these templates, a description of the template-to-QTI implementation is provided. In addition, a set of exemplar items is provided. These exemplars **are** aligned to specific targets and have been approved by the Consortium. For each exemplar, the item sample specification is provided along with a movie depicting the student interaction. Some additional exemplar items are included from templates that were not selected in this initial operational stage.

Choosing a Template

To produce evidence about a claim, the students must take some action. Decide what action you want the student to take and then use the template that captures that action.

I want the student to...	Use TE ITEM Template	Notes
Categorize or classify objects (short text and/or graphics)	Select and Order or Classification or Tiling	<ul style="list-style-type: none"> • templates support classification of graphics or short pieces of text (such as 1-2 word phrases) • use Select and Order if only <i>one</i> object is allowed in each category <i>and</i> the number of objects is pre-defined by item writer • use Classification if <i>more than one</i> object is allowed in each category <i>and</i> the number of objects is pre-defined by item writer • use Tiling if the objects are pre-defined by the item-writer, but the student can instantiate (i.e., spawn) multiple copies of each object; Tiling can support one or more object per category
Connect objects (text and/or graphic) together (e.g., matching)	Connections	
Create a bar chart from data	Select Defined Partitions	<ul style="list-style-type: none"> • the data is pre-specified by the item writer for the student (in tabular, textual, or other format)
Create a dot plot	Placing Points	
Create a line plot	Partition Line Then Place Points	<ul style="list-style-type: none"> • template supports items asking for Xs above numeric categories
Draw a circle	Single Circle	
Draw a line and choose one half of the plane created by the line	Single Line Then Select	
Draw a parabola to match an equation	Single Parabola	
Draw a polygon with certain characteristics	Vertex-Based Triangles or Vertex-Based Quadrilaterals	<ul style="list-style-type: none"> • use Vertex-Based Triangles if the correct response is a triangle (note that the student might create a response with a polygon that is not a triangle) • use Vertex-Based Quadrilaterals if the correct response is a quadrilateral (note that the student might create a response with a polygon that is not a quadrilateral)

I want the student to...	Use TE ITEM Template	Notes
Draw a polygon that is a transformation (rotation, reflection, translation, or dilation) of an existing polygon	Vertex-Based Triangles or Vertex-Based Quadrilaterals	<ul style="list-style-type: none"> • use Vertex-Based Triangles if the correct response is a triangle (note that the student might create a response with a polygon that is not a triangle) • use Vertex-Based Quadrilaterals if the correct response is a quadrilateral (note that the student might create a response with a polygon that is not a quadrilateral)
Draw a ray	Single Ray	
Draw an angle	Single Angle	
Draw line(s) that match equations	Single Lines or Straight Lines	<ul style="list-style-type: none"> • use Single Lines if you want the student draw one line (and the correct response is a single line) • use Straight Lines if the correct response is two or more lines (note that the student might create a response with only one line to items generated from the Straight Lines template)
Draw line segment(s)	Single Lines or Straight Lines	<ul style="list-style-type: none"> • use Single Lines if you want the student draw one line segment (and the correct response is a single line segment) • use Straight Lines if the correct response is two or more line segments (note that the student might create a response with only one line segment to items generated from the Straight Lines template)
Draw multiple lines and choose an intersection of the lines	Multiple Lines Then Select	
Draw parallel, perpendicular, or intersecting lines	Straight Lines	
Extend a pattern	Tiling	
Fill in the blanks with objects (text and/or graphics)	Select and Order or Tiling	<ul style="list-style-type: none"> • use Select and Order if both the objects and the number of objects is pre-defined by the item writer • use Tiling if the objects are pre-defined by the item-writer, but the student can instantiate (i.e., spawn) multiple copies of each object

I want the student to...	Use TE ITEM Template	Notes
Generate a number sentence	Select and Order or Tiling	<ul style="list-style-type: none"> • both of these templates can only be used to create items where the potential components of the number sentence are pre-defined; this template does not support students typing in number sentences • use Select and Order if the components and the number of each component that can be used is pre-defined by item writer • use Tiling if the components are pre-defined by the item-writer, but the student can instantiate (i.e., spawn) multiple copies of each component
Generate an equation	Select and Order or Tiling	<ul style="list-style-type: none"> • both of these templates can only be used to create items where the potential components of the equation are pre-defined; his template does not support students typing in equations • use Select and Order if the components and the number of each component that can be used is pre-defined by item writer • use Tiling if the components are pre-defined by the item-writer, but the student can instantiate (i.e., spawn) multiple copies of each component
Generate a pattern	Tiling	<ul style="list-style-type: none"> • the item writer must be able to specify all possible patterns that are considered correct responses
Identify one or more coordinate points	Placing points	
Identify one or more locations on a graphic	Placing Points	
Identify points on one or more number lines	Selecting Points and Ranges on Number Lines	
Identify sections or ranges of one or more number lines	Selecting Points and Ranges on Number Lines	

I want the student to...	Use TE ITEM Template	Notes
Order objects (short text and/or graphics)	Select and Order	<ul style="list-style-type: none"> • templates support ordering of graphics or short pieces of text (such as 1-2 word phrases) • in items generated from this template, the objects and number of “ordering slots” are pre-defined by the item writer
Match objects (text and/or graphic) to other objects	Connections	
Partition a figure and select some of the partitions	Partition Object Then Select	<ul style="list-style-type: none"> • the figure can be a rectangle or a circle
Partition a number line and place points or markers at places along the line	Partition Line Then Place Points	
Plot points on a grid	Placing Points	
Plot points on a number line	Selecting Points and Ranges on Number Lines	
Put objects (short text and/or graphics) in order	Select and Order	<ul style="list-style-type: none"> • templates support ordering of graphics or short pieces of text (such as 1-2 word phrases) • in items generated from this template, the objects and number of “ordering slots” are pre-defined by the item writer
Put objects (short text and/or graphics) into categories	Select and Order or Classification or Tiling	<ul style="list-style-type: none"> • templates support classification of graphics or short pieces of text (such as 1-2 word phrases) • use Select and Order if only one object is allowed in each category and the number of objects is pre-defined by item writer • use Classification if more than one object is allowed in each category and the number of objects is pre-defined by item writer • use Tiling if the objects are pre-defined by the item-writer, but the student can instantiate (i.e., spawn) multiple copies of each object; Tiling can support one or more object per category
Rearrange the order of text (words, phrases, sentences, or paragraphs)	Reorder Text	<ul style="list-style-type: none"> • the item writer specifies the “chunks” or divisions of text that will be rearranged

I want the student to...	Use TE ITEM Template	Notes
Select a single piece of text and type in a replacement for that text	Select and Change Text	<ul style="list-style-type: none"> the item writer must list all possible correct keys for the replaced text
Select one or more objects	Select Objects	<ul style="list-style-type: none"> objects can contain graphics, text, or both
Select one or more pieces of text as their response (words, phrases, sentences, or paragraphs)	Select Text	<ul style="list-style-type: none"> the item writer specifies the “size of the chunks” or divisions of text (e.g., only whole words can be selected, only whole sentences can be selected, etc.)
Select options from drop-down menus	Drop Downs	
Select pre-defined partitions of a figure	Select Defined Partitions	<ul style="list-style-type: none"> the figure can be rectangle, a circle, or a bar the item writer defines the number of figures and the partitions of each figure
Transform a figure (translate, rotate, reflect, or dilate)	Object Transform	<ul style="list-style-type: none"> the figure must be pre-defined by the item writer the student does not draw a new figure, but transforms the figure provided

TE Items Common Terms and Definitions

Artificial Intelligence (AI) Task: An AI task is composed of multiple items for which there is a dependency between responses that requires advanced programming logic. For example, in an AI task composed of three items, the response to the third item may be dependent on the response to the second item. These items require custom programming and AI-based scoring. As a result, they are more expensive to develop, replicate, and bring to scale. (This is distinct from adaptive testing, which bases the *selection* of items on previous responses.) For the current project, AI tasks are not included for consideration directly as template-based items. AI tasks may be considered through the adoption of scoring engines, as described in the Consortium’s statement on principles for adoption of AI.

Boolean: A binary variable with two possible values: true and false. These variables can be used to indicate the preference for any characteristic that has two states, e.g., on/off or yes/no. For example, whether a default coordinate grid should be displayed on an item can be specified by the item writer as yes (true) or no (false). Whether a student is limited to choosing one item or is allowed to choose multiple items can be specified by the item writer as limited (true) or unlimited (false).

Composite Set (of Items): A composite set is composed of multiple individual items (either technology-enhanced items, technology-enabled items, or traditional items), with each item containing its own interaction, score response set, and scoring algorithm. The items within the composite set are presented to the student together. Each individual item returns an individual score that is independently calculated (i.e., the scoring of one item response is not dependent on the response to a prior item). The score of the composite set is based on a combination of the scores from each independent composition. To determine the composite set score, a composite set includes a scoring algorithm that defines how scores from the multiple individual items are combined to provide a single composite set score.

Dichotomous Item: An item whose responses are scored as correct or incorrect.

Interaction: An interaction is the action that a student performs to produce a response. A small set of examples of common interactions include:

1. Select a single option from a set of options (traditional selected-response)
2. Select multiple options from among a set of options
3. Enter a text string (traditional constructed-response)
4. Create a line
5. Move one or more objects to given set of locations (drag-and-drop)
6. Produce a geometric shape
7. Select blocks of text
8. Mark a visual field

There are many more interactions that a student can perform to produce a response; these are provided only as a small set of examples.

Interaction Instructions: Interaction Instructions are provided in each template specification as a detailed description of how the student would interact with an item developed from the template. For example, “To move a point, click the point to select it, move the pointer, and click again to place the point.” The interaction instructions may need to be modified based on the assessment delivery system (e.g., click is inappropriate for tablets, some keys inappropriate for Mac/Windows, etc.) or on grade level appropriateness.

Interaction instructions are intended to inform the training of students prior to the assessment. These instructions should be available to students during item presentation upon request, but not included in the layout of the item. A button to reveal or hide interaction instructions near the appropriate interaction space is recommended as best-practice.

Desired-response instructions are straightforward directions for the action the student is to take (without necessarily specifying *how* to accomplish that action). For example, “Draw a line...” or “Choose the sentence that...” These item directions should be included in the item prompt.

Assessment delivery systems will need a thorough understanding of when the appropriate directions/instructions should be delivered to the student given the assessment context.

Interaction Space Parameter: The interaction space is the area within the item where the interaction occurs. Interaction space parameters define characteristics of this area, e.g., whether a grid should be shown, whether the number of objects that can be selected should be limited, etc.

Item: An assessment question that contains one or more interactions with response data collected as a result of the interaction. An item should contain its own, unique prompt. It could have its own stimulus, and/or share a stimulus with other items.

Item Cluster: An item cluster is a group of items that share a stimulus. Items in a cluster can return a cluster-based score for use in adaptive testing. For example, a collection of reading comprehension items that share a passage might be an item cluster. As another example, a mathematics scenario is presented such as budgeting for a vacation and then multiple items are asked about this scenario. The adaptive engine will make a judgment based on the collective score of the items in the cluster.

Polytomous Item: An item whose responses are scored on a scale, e.g., 0-4 or 0-2.

Response Data: Response data is the information that is returned by the item as a result of the student's interaction. For selected-response, response data is the answer option selected by the student. For constructed-response, response data is the text string produced by the student. For a "create a line" item, the response data is the function of the line. For a "move object" item, the response data is the location of the object. What is contained in response data is specific to each template and is sufficient to implement the scoring rule to provide a score.

Response Reversal: Templates provide two methods for students to reverse their actions: a Reset button and a Clear button. A Reset button will be provided in templates that require the student to create a single piece of content (e.g., Single Line, Single Angle, etc.). Clicking the Reset button will result in the removal of the content created. A Reset button will also be provided in templates that do not require the student to create content, but only to manipulate content provided in the item (e.g., Reorder Text, Classification, etc.). Clicking the Reset button will result in the content returning to its original position or state. Confirmation will be required before the content is reset.

A Clear button will be provided in templates that require the student to create multiple pieces of content (e.g., Placing Points, Straight Lines, etc.). Students will click a single object to select the object, then click the Clear button to remove the object. In these templates, a Reset button will also be included that will remove *all* objects. Confirmation will be required before content is reset.

Some templates ask students to select content (e.g., Select Text, Select Defined Partitions, etc.). In these templates, clicking on a selected object will de-select, or clear, the selection. This is not considered part of the formal Response Reversal options provided, but part of the interaction itself. Thus, for these templates, a Reset and/or Clear button will also be provided.

Score Response Set: A score response set provides information that is used to assign a score to a response produced through an interaction. The score response set may contain a variety of information depending on the interaction type and the evidence that is collected through that interaction. This is similar to a *key*.

For example, for a selected-response item, the score response set contains the correct response. For a selected-response item with multiple responses, the score response set contains each of the correct response options. The score response set might also contain a subset of correct response options associated with a partial score (defined by the item writer).

As another example, for a "draw a line" item, the score response set may contain the slope and y-intercept that represents the correct response. The score response set may also contain information about allowable tolerance for the slope and/or the y-intercept. In addition, the score response data might contain Boolean values that specify whether to score the response data based on comparison of slope only, y-intercept only, or both. In the case that there are multiple correct responses (e.g., for an item asking a student to create a line of symmetry for a square), the score response set contains information about each correct response.

Finally, a score response set includes information about the number of score points that may be awarded for an item (e.g., dichotomous or polytomous). For polytomous scores, response data is provided for each possible score value by the item writer.

Scoring Algorithm: A scoring algorithm defines the logic applied to compare response data to pre-defined score response sets and to award a specified score when the response data matches a given score response set. For a single option selection item (e.g., traditional selected-response), the scoring algorithm specifies that a score of 1 is awarded when the student response data matches the score response set (i.e., the correct response). For a “draw a line” item, the scoring algorithm might compare the slope of the response data to the slope of the line described in the score response set. Alternatively, the scoring algorithm might consider both the slope and the y-intercept. Scoring algorithms can be complex and consider multiple score response sets. Scoring algorithms can be specified to yield dichotomous, polytomous, or partial credit (e.g., 0, 0.5, or 1 out of 1) scores.

Tab-Enter Navigation: the ability to use two basic commands on a computer, namely the TAB key command and the ENTER key command. Users can navigate to an object using TAB, then execute the function of the object (like a button) using ENTER. Those commands can be assigned to basic input devices, only two examples of which are large buttons and sip and puff devices (technology used to send signals to a device using air pressure by inhaling or exhaling on a straw, tube or "wand;" primarily used by people who do not have the use of their hands).

Technology Enabled Items (Constructed-Response or Selected-Response): Computer delivered items that use digital media as the stimulus (sound, video, or interactive widget), but do not require specialized interactions for response and/or accompanying response data. Technology Enabled Items also allow for non-traditional layout of items that use constructed-response and/or selected-responses.

Technology Enhanced Items (TEI): Computer delivered items that include specialized interactions for response and/or accompanying response data. These include interactions/responses that are not selected-response or text-entry. TEI may include digital media as the stimulus (e.g., sound, video, interactive widget, etc.).

TEI Template: A TEI template describes a single interaction, response data collected as a result of that interaction, and the logic applied to score response data. Templates can support dichotomous and polytomous scoring. TEI templates can be used across content areas, but are sometimes better suited for one content area. For example, classifying objects is well suited for both Math and ELA. Selecting text can be used to create an item in any content area, but is likely more appropriate for ELA items. Drawing a parabola could, in theory, be used to create an item in any content area, but this interaction is most likely appropriate for Math items.

Innovations in the prompt or stimulus through new media, such as full-motion animation, may be included in TEI. These additions are not exemplified in the template specifications as they do not require new interactions or formats of response data, as described in the templates.

Tolerance: A tolerance is an acceptable interval around an expected value. Tolerances account for difficulties or imprecision that result from the test-taking interface rather than the student’s cognitive abilities. Tolerances should be determined through field testing and are not included in the template specifications.

Literature Review

Traditional selected-response (SR) items are frequently inadequate for measuring high-level knowledge and understanding (Archbald & Newmann, 1988; Birenbaum & Tatsuoka, 1987; Darling-Hammond & Lieberman, 1992). One solution to the shortcomings of SR items is the use of constructed-response (CR) items. These items are frequently used to measure high-order skills and knowledge, but are subjective, timely, and expensive to score. In recent years, researchers have leveraged technological advancements to combine the measurement power of CR items with the automated-scoring capability of SR. Some research has focused on automated text and essay scoring (e.g., Dikli, 2006) and automated scoring of mathematical graphs and expressions (Bennett, Morley, & Quardt, 2000a, 200b).

Another branch of research has focused on the development and evaluation of technology-enhanced (TE) items. TE items provide specialized interactions that allow students to interact with digital content while still supporting automated scoring (Parshall, Davey, & Pashley, 2000; Zenisky & Sireci, 2002). TE items are grouped based on the interaction that the student uses to respond to the item, e.g., dragging-and-dropping, sorting, ordering, editing, and highlighting, etc. No exhaustive list is available because rapidly developing technologies frequently support new interactions, creating a single line, partitioning a rectangle, etc. TE items can use parameters to define aspects of student responses that contribute evidence of the claim and thus should contribute to the scoring of the response. For example, in a sorting item, the different orderings of content and the associated score of each ordering is pre-defined in the item's parameters. In an item asking students to draw a line, the slope and y-intercept of different lines and the associated scores of those lines are pre-defined in the item's parameters.

TE items offer many benefits over traditional items. The most significant advantage is that TE items can provide improved measurement of certain constructs, specifically high-level constructs, often in a more realistic context (Bennett, 1999; Gorin, 2006; Huff & Sireci, 2001; Jodoin, 2003; Sireci & Zenisky, 2006; Zenisky & Sireci, 2002). A second advantage of TE items is that they reduce the effect of random guessing (Huff & Sireci, 2001). CR items have always reduced the impact of test-taking skills and guessing, but TE items allow these benefits to be leveraged on items administered via computer and automatically scored. TE items frequently require students to *produce* information, rather than simply *select* information, which is often a more authentic form of measurement (Archbald & Newmann, 1988). A third advantage of using TE items is the reduction of cognitive load from non-relevant constructs, such as the reading load for non-reading related items or the cognitive load required to keep various item constructs in memory (Kumar, White, & Helgeson, 1993). Finally, TE items tend to be more engaging to students, a significant advantage in an era when students frequently feel over-tested (Strain-Seymour, Way, & Dolan, 2009; Dolan, Goodman, Strain-Seymour, Adams, & Sethuraman, 2011).

- Archbald, D. A. & Newmann, F. M. (1988). *Beyond Standardized Testing: Assessing Authentic Academic Achievement in Secondary School*. Reston, VA: National Association of Secondary School Principals.
- Bennett, R. E. (1999). Using new technology to improve assessment. *Educational Measurement: Issues and Practice*, 18(3), 5-12.
- Bennett, R. E., Morley M., & Quardt, D. (2000a). Graphical modeling: A new response type for measuring the qualitative component of mathematical reasoning. *Applied Measurement in Education*, 13, 303-322.
- Bennett, R. E., Morley M., & Quardt, D. (2000b). Three response types for broadening the conception of mathematical problem solving in computerized tests. *Applied Psychological Measurement*, 24(4), 294-309.
- Birenbaum, M., Kelly, A. E. & Tatsuoka, K. (1992). Towards a stable diagnostic representation of students' errors in algebra. ETS-RR-92-58-ONR. ERIC Clearinghouse Document ED 356973.
- Darling-Hammond, L. & Lieberman, A. (1992). The shortcomings of standardized tests. *The Chronicle of Higher Education*, B1-B2.
- Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning, and Assessment*, 5(1).
- Gorin, J. S. (2006). Test design with cognition in mind. *Educational Measurement: Issues and Practice*, 25(4), 21-35.
- Huff, K. L. & Sireci, S. G. (2001). Validity issues in computer-based testing. *Educational Measurement: Issues and Practice*. 20(3), 16-25.
- Jodoin, M. G. (2003). Measurement efficiency of innovative item formats in computer-based testing. *Journal of Educational Measurement*, 40(1), 1-15.
- Kumar, D., White, A. L., & Helgeson, S. L. (1993). Effect of HyperCard and traditional performance assessment methods on expert-novice chemistry problem solving. Paper presented at the *Annual Meeting of the National Association for Research in Science Teaching*. Atlanta, GA.
- Parshall, C. G., Davey, T., & Pashley, P. J. (2002). Innovating item types for computerized testing. In W. J. van der Linden and C. A. W. Glas (Eds.), *Computerized Adaptive Testing: Theory and Practice* (pp.129-148).
- Sireci, S. G. & Zenisky, A. L. (2006). Innovative item formats in computer-based testing: In pursuit of improved construct representation. In S. M. Downing and T. M. Haladyna (Eds.), *Handbook of test development* (pp. 329-347). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Strain-Seymour, E., Way, W., & Dolan, R. (2009). *Strategies and Processes for Developing Innovative Items in Large-Scale Assessments*. Iowa City, IA: Pearson Education.
- Dolan, R. P., Goodman, J., Strain-Seymour, E., Adams, J., & Sethuraman, S. (2011). *Cognitive Lab Evaluation of Innovative Items in Mathematics and English Language Arts Assessment of Elementary, Middle, and High School Students*. Iowa City, IA: Pearson Education.
- Zenisky, A. L. & Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education*, 15(4), 337-362.