

Chapter 3: Item Development.....	2
Item and Task Specifications	2
The Item/task Pool.....	4
Item Writing	4
Item Reviews	6
Item Scoring	7
Composition of the Item Pool.....	8
Table 1 Item Types found in the Consortium’s Item Pool.....	8
Table 2 Total Number of CAT Items and Performance Tasks (PT) developed by Grade and Content Area.....	10
Table 3. Mathematics Specifications and Archetype Delivery	11
Table 4. English/Language Arts/Literacy Specifications and Archetype Delivery	12
Use of Systems and Tagging	12
Training Activities	13
Virtual Training of Educator Item and Task Writers	14
Continued Training during Development.....	14
Certification and Management	14
Table 5. Sample Certification Set	15
Item/Task Set Evaluation	15
Role of Item Quality Review Panel	15
Educator Recruitment Activities.....	16
Table 6. Recruiting Activities and Timeline	17
Item development process.....	17
Table 7. Additional item writing, development, review and scoring documentation	20

Chapter 3: Item Development

In order to build a summative assessment that measured the intended claims, the Consortium's test development cycle was iterative, involving experts from various education-related fields, and was based on assessment-related research and best practices.

Item and Task Specifications¹

The item specifications bridge the span from the content specifications and ALDs to the assessment itself. While the content specifications established the Consortium's claims and the types of evidence or targets, that would need to be collected in order to support these claims, more specificity was needed in order to develop items and tasks that measured the claims. Working with three vendors (ETS, Measured Progress, and CTB) with extensive experience in item writing, the Consortium's Item Development Work Group and Performance Task Work Group created item and task specifications for both ELA/Literacy and mathematics.

The first iteration of the item and task specifications were developed in 2011. In early 2012, the Consortium held a series of showcases where the contractors introduced the item and task specifications and collected feedback from member states. Using this feedback, the item and task specifications were revised during the first quarter of 2012.

Using the revised item and task specifications, a small set of items was developed and administered in fall 2012 during a small-scale trial. This provided the Consortium with their first opportunity to administer and score the new item types. During the small-scale trials, the Consortium also conducted cognitive laboratories to better understand how students respond to various types of items. The cognitive laboratories used a think-aloud methodology in which students speak their thoughts while working on a test item. The item and task specifications were again revised based on the findings of the cognitive laboratories and the small-scale trial. These revised specifications were used to develop items for the 2013 pilot test, and they were again revised based on 2013 pilot test results and subsequent review by content experts.

The Consortium's item and task specifications are designed to ensure that the assessment items measure the assessment's claims. Indeed, the purpose of the item and task specifications is to define the characteristics of the items and tasks that will provide the evidence to support one or more claims. To do this, the item and task specifications delineate the types of evidence, or targets, that should be elicited for each claim within a grade level. Then, they provide explicit guidance on how to write items in order to elicit the desired evidence.

In doing this, the item and task specifications provide guidance on how to measure the targets (standards) first found in the content specifications. The item and task specifications provide guidelines on how to create the items that are specific to each assessment target and claim through

¹ <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/ItemSpecifications/GeneralItemSpecifications.pdf>

the use of task models. In mathematics a task model provides a description of an item/task's key features. These task models describe the knowledge, skills, and processes being measured by each of the item types aligned to particular targets. In addition, the task models sometimes provide examples of plausible distractors. Exemplar items are provided within every task model. In ELA these functions are carried out through item specifications.

These task models were developed for each grade level and target in order to delineate the expectations of knowledge and skill to be included on test questions in each grade. In addition, both the ELA/L and mathematics item and stimulus specifications provide guidance on determining the grade appropriateness of task and stimulus materials (the materials that a student must refer to in working on a test question). The task and stimulus models also provide information on characteristics of stimuli or activities to avoid because they are not important to the knowledge, skill, or process being measured.

This is important because it underscores the Consortium's efforts to develop items that are accessible to the widest range of students possible; in other words, Consortium items are created according to the principle of **universal design**. As the name suggests, the concept of universal design aims to create items that accurately measure the assessment target for all students. At the same time, universal design recognizes that one solution rarely works for all students. Instead, this framework acknowledges "the need for alternatives to suit many different people."²

To facilitate the application of universal design principles, item writers are trained to consider the full range of students who may answer a test question. A simple example of this is the use of vocabulary that is expected to be known by all third-grade students versus only those third-grade students who play basketball. Almost all third-grade students are familiar with activities (e.g., recess) that happen during their school day, while only a subset of these students will be familiar with basketball terms like "double dribble," "layup," "zone defense," or "full-court press."

Classroom Activities

In order to mitigate possible unfamiliarity with context or vocabulary performance tasks are preceded by an unscored activity conducted with the whole classroom. The activity is intended to resolve issues of unfamiliarity so that performance on the task reflects only subject matter knowledge. There are several performance tasks associated with each classroom activity so that the class receives a variety of tasks.

In addition to this, the item specifications discuss accessibility issues that are unique to the creation of items for a particular claim and/or assessment target. The accessibility concerns discuss the different supports that various groups of students may need to access the content of an item. By

² Rose, D. & Meyer, A. (2000). Universal design for learning, associate editor column. *Journal of Special Education Technology* 15(1):66-67

considering the possible supports that may be needed for each item, item writers are able to create items that will support those adaptations.

The use of universal design principles allows the Consortium to collect evidence on the widest possible range of students. By writing items that adhere to the item and task specifications, the Consortium is assured that the assessments measure the claims and assessment targets established in the content specifications as well as the knowledge, skills, and processes found in the Common Core State Standards for *all* students for whom the assessment is appropriate.

The Item/task Pool

An **item pool** refers to a collection of test questions (known as items) measuring the same content area (e.g., mathematics) within the same grade. (As explained in the Test Design chapter, the use of off-grade-level items is allowed in some instances.) The quality of the items is a primary concern when building an item pool. The Consortium took multiple steps to ensure the quality of the items in our item pool. Building on the ongoing process of developing item/task specifications and test blueprints described in the previous chapter, the Consortium used an iterative process for creating and revising each item as well as the collection of items. The Consortium tested items and refined its approach to item development through three steps: small-scale tryouts, a large pilot test, and the largest ever field test of a K-12 assessment. Details of the pilot and field tests are found in CH _____. During each phase, the Consortium used cognitive laboratories to understand the strategies that students used to respond to the items. By incorporating this tiered and iterative approach, the item and task specifications that guided the development of the final operational pool were improved based on the lessons learned during these important tryouts.

Using the summative and comprehensive interim test blueprints, the number and distribution of items to be written were specified for item writing teams. Pools of items/tasks were written specifically to support the operational blueprint. Teachers were integrally involved in the creation of the item/task pool from beginning to end. Some participated in the processes described in the flow charts below. Others developed many of our items through a rigorous item writing process, and yet others reviewed the items for accuracy and appropriateness of the content knowledge and skill level required to respond to the items, potential issues of bias in favor of or against any demographic group of students, and accessibility for students with disabilities and English language learners. Teams of educators reviewed items for content, bias, and accessibility prior to administration to any students. Following the pilot and field test administrations, items were again reviewed if pilot or field test data indicated a potential problem. Finally, teachers participated in in range finding and scoring constructed-response items/tasks to ensure that the items/tasks could be properly scored given their scoring rubrics.

In this section, we will examine the primary role that educators played in creating the field-test item pool by writing, reviewing, and scoring items. This section will end by examining the current composition of the item pool.

Item Writing

The job of writing all of the items and performance tasks was no small undertaking, and the Consortium worked with educators throughout the test development cycle to develop items. Prior to the spring 2013 pilot test, the Consortium engaged 136 educators in K-12 and higher education

from 19 member states to write items. Prior to the spring 2014 field test, 184 educators in K-12 and higher education from 16 member states participated in item writing. All K-12 participants:

- Were certified/licensed to teach ELA/L and/or mathematics in a K-12 public school;
- Were currently teaching in a public school within a Smarter Balanced Governing State;
- Had taught ELA and/or mathematics in grades 3 through 8 and/or high school within the past three years (second-grade teachers were also recruited to participate in the development of grade 3 items and/or tasks);
- Had previously reviewed part or all of the CCSS for the content area for which they were writing items and/or performance tasks;
- Submitted a statement of interest that described their interest in developing Smarter Balanced items and/or performance tasks as well as their qualifications for doing so;
- Completed training and achieved qualifications through the certification process.

Qualifications for Higher Education Faculty included:

- Current employment with, or recent retirement from, a college or university located within a Smarter Balanced Governing State;
- Having taught developmental and/or entry-level courses in English, composition, mathematics, statistics or a related discipline within the last 3 years;
- Having previously reviewed part or all of the CCSS for the content area in which they are interested in writing items and/or performance tasks;
- Completing training and achieving qualifications through the certification process.

The selected educators were trained on the Consortium's content specifications, the item and task specifications, and stimulus specifications (ELA/L) as well as the item authoring system in which the items were developed. In addition, professional item writers and the Consortium held regular meetings to provide direction and feedback to the educators. Educators, state partners, and assessment vendors developed the items in the Consortium's item pool.

Training

Educators participated in a series of facilitated, online webinars in order to qualify as item writers. To facilitate participation, the Consortium scheduled multiple sessions in different time zones, including evening sessions. In addition to the facilitated sessions, the Consortium provided training modules that covered background on the Consortium, assessment design principles, and detailed information about item and performance task development. All modules were available in three formats: a PowerPoint presentation with notes, a streaming presentation with narration that could be viewed online, and a downloadable audio/video presentation.

The item writers were specifically trained on the Consortium's content and item specifications, stimulus specifications,³ sensitivity and bias guidelines,⁴ and general accessibility guidelines.⁵ Training on these specifications and guidelines helped ensure that item writers were trained to write items that allowed the widest possible range of students to demonstrate their knowledge, skills, and cognitive processes in regard to the content. This meant that item writers needed to understand the content for which they were writing items as well as accessibility and sensitivity issues that might hinder students' ability to answer an item. Item writers were also trained to be aware of issues that might unintentionally bias an item for or against a particular group.

Educator Participation

Consistent with the Consortium process, educators were the primary developers of items. The active involvement of educators was critical to the success of the item writing activities. Educators engage with students on a daily basis, and they understand the ways in which students can demonstrate their knowledge. Their involvement in item writing helped ensure that the assessment system is accurate and efficient, and provides valid evidence of student learning.

State-Managed Item Development

The Consortium invited member states to participate in a separate effort to write items. This voluntary effort, known as State-Managed Item Development, was conducted to build the capacity of states to write items and to support the overall sustainability of the Consortium. To this end, six states (HI, ID, MI, WA, WV, and WY) participated in the state-managed field test item development opportunity. During this opportunity, educators within the six states developed approximately 3,100 items in mathematics and ELA/L across grades 3 through 8 and high school. Many of these items were field tested during the operational test in spring 2015.

Item Reviews

Once items were written, groups of educators reviewed items prior to their pilot test administration in spring 2013 and their field test administration in spring 2014. Items that survived the pilot test were again reviewed prior to their use in the spring 2014 field test.

Accessibility, Bias/Sensitivity, and Content Reviews

Panels of educators reviewed all items, performance tasks, and item stimuli for accessibility, bias/sensitivity, and content. (Item stimuli refer to the reading passages used on the ELA/L assessments or the figures and graphics used on the mathematics assessments.) Prior to the spring 2013 field test, 122 ELA/L educators and 106 mathematics educators reviewed items and performance tasks for accessibility, bias/sensitivity, or content, and 60 educators reviewed the

³ <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/EnglishLanguageArtsLiteracy/ELAStimulusSpecifications.pdf>

⁴ <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/Guidelines/BiasandSensitivity/BiasandSensitivityGuidelines.pdf>

⁵ <http://www.smarterbalanced.org/wordpress/wp-content/uploads/2012/05/TaskItemSpecifications/Guidelines/AccessibilityandAccommodations/GeneralAccessibilityGuidelines.pdf>

ELA/L stimuli. Prior to the spring 2014 field test, 107 ELA/L educators and 157 mathematics educators from 14 states reviewed items and performance, and 95 educators from 13 states reviewed the ELA/L stimuli.

The educator qualifications for the accessibility, bias/sensitivity, and content reviews were the same as the educator qualifications for item writing except that participants were not required to submit a statement of interest. In addition, it was preferred (but not required) that educators have previous experience reviewing items, tasks, and/or stimuli.

During the accessibility reviews, panelists identified issues that may negatively affect a student's ability to access stimuli, items, or performance tasks, or to elicit valid evidence about an assessment target. During the bias and sensitivity review, panelists identified content in stimuli, items, or performance tasks that may negatively affect a student's ability to produce a correct response because of their background. The content review focused on developmental appropriateness and alignment of stimuli, items, and tasks to the content specifications and appropriate depths of knowledge. Panelists in the content review also checked the accuracy of the content, answer keys, and scoring materials. Items flagged for accessibility, bias/sensitivity, and/or content concerns were either revised to address the issues identified by the panelists or removed from the item pool.

Data-Related Reviews

The items developed for the item pool were administered during the spring 2013 and spring 2014 pilot tests, and the pilot test data from both administrations were analyzed to examine the statistical quality of the items in the pool. The Consortium established statistical criteria to flag items for possible defects in quality related to content, bias, or accessibility. For example, content-related criteria flagged items for further review if they were extremely difficult or extremely easy. Accessibility-related criteria flagged items that were differentially more difficult for students with disabilities compared to students without disabilities.

Following the spring 2013 pilot, 40 educators participated in the item data review and examined the items for possible content-related issues or accessibility-related issues. Following the spring 2014 pilot, 57 ELA/L educators from 16 states and 30 mathematics educators from 12 states participated in item data review, examining the items for possible content-related issues or accessibility-related issues. At least two educators reviewed each item. These educators were trained via webinars on the flagging criteria and on how to evaluate flagged items. These educators made recommendations on whether to accept the item with no change, revise and re-field test the item, or reject the item from the pool. McGraw-Hill CTB content experts reviewed all items where the reviewers' recommendations disagreed. In addition, McGraw-Hill CTB content experts and psychometricians also reviewed and provided recommendations for all items where both reviewers recommended accepting the item. In each situation, the content expert provided the Consortium with a final recommendation for the item.

The educator qualifications for the item data reviews were the same as the educator qualifications for item writing except that participants were not required to submit a statement of interest.

Item Scoring

For those items that could not be machine scored, the Consortium engaged 102 participants from 20 states in range finding activities for those items requiring human scoring following the spring

2013 pilot. After the spring 2014 pilot, 104 educators participated in range finding. Range finding improves the consistency and validity of scoring for the assessment. During range finding, the educators focused on the performance tasks for mathematics and ELA/L. In mathematics, educators also reviewed constructed response items for grades 7, 8, and high school. During range finding, the participants reviewed student responses against item rubrics, validated the rubrics’ effectiveness, and selected the anchor papers that would be used by professional scorers during the main scoring event.

The educator qualifications for range finding were the same as the educator qualifications for item writing, except that participants were not required to submit a statement of interest. In addition, it was preferred (but not required) that educators had previous range finding experience.

Composition of the Item Pool⁶

The Consortium developed many different types of items beyond the traditional multiple-choice item. This was done to measure the claims and assessment targets with varying degrees of complexity by allowing students to construct their responses rather than simply recognizing a correct response. These different item types are listed in Table 1 below.

Table 1 Item Types found in the Consortium’s Item Pool.

Item Types	ELA/L	Mathematics
Multiple Choice, Single Correct Response	X	X
Multiple Choice, Multiple Correct Response	X	X
Two-part Multiple Choice, with Evidence Responses (EBSR)	X	
Matching Tables	X	X
Hot Text	X	X
Drag and Drop	X	X
Short Text Response	X	
Essay	X	
Hot Spot		X
Short Text and Fill-in Tables		X

⁶ Examples of many of the item types may be found at: <http://www.smarterbalanced.org/sample-items-and-performance-tasks/>.

Each grade's item pool for the Consortium's test was necessarily large to support the summative and interim assessments⁷ being delivered via a computer using adaptive test-delivery technology, commonly called a computer adaptive test or CAT. Unlike a traditional paper-and-pencil test where all students take the same items, students taking the Consortium's CAT will take items and tasks targeted to their ability level. This means that the Consortium needed to develop a very large number of items in order to meet the needs of the student population.

In addition to the items for the CAT, the Consortium also developed performance tasks. All students take performance tasks that are designed to measure a student's ability to integrate knowledge and skills across multiple assessment targets. These performance tasks may also be delivered via the same online assessment delivery system as the CAT.

Table 2 below shows the total number of CAT items and performance tasks (PT) found in each item pool by grade level and content area. As the table shows, over 1,600 ELA/L CAT items were developed in each of grades 3 – 8, and 5,711 items were developed for high school. In mathematics, approximately 1,500 items were developed in each of grades 3 – 8, and 4,512 items were developed for high school. The items in these pools will support both the summative and interim assessments.

There were approximately 50 PTs per grade developed in each of grades 3 – 8 in both ELA/L and mathematics. In high school, the Consortium created 124 ELA/L PTs and 132 mathematics PTs. Each PT has multiple associated items: four and six items per PT in ELA/L and mathematics, respectively.

⁷ Interim assessments will not be delivered via CAT until the 2015-16 school year. They are fixed-form tests in the 2014-15 school year.

Table 2 Total Number of CAT Items and Performance Tasks (PT) developed by Grade and Content Area

Row Labels	ELA/L		Math	
	CAT	PT	CAT	PT
3	1711	49	1502	50
4	1653	49	1551	49
5	1659	47	1517	49
6	1683	47	1503	49
7	1657	47	1487	49
8	1626	49	1488	49
HS	5711	124	4512	132
Grand Total	15700	412	13560	427

The numbers of items that survived to be field tested is listed in Chapter 7.

Selection of Items for the Operational Item Pool

The statistical quality of the items was again evaluated following the 2014 field test. Items that did not perform well according to established psychometric criteria (for example, item statistics such as difficulty and discrimination) were forwarded to content experts for review. The same psychometric criteria were used to judge items regardless of whether the items were used on the interim assessment or the summative assessment.

For the first operational year (2014-2015), items for both the interim assessment and the summative assessment were drawn from the same item pool. The summative item pool is secure, while the interim pool can be accessed by teachers to aid in planning and interpretation. The long-term plan is that most items will first be administered on the summative assessment before entering the interim item pool. In the first operational year, interim pools supported fixed form tests. Many of the items being field tested in 2015 will move directly from the field test to the interim item pool as necessary to meet the content requirements and support interim adaptive testing where possible.

Table 3. Mathematics Specifications and Archetype Delivery

MATH						
Delivery Number	Number of Specs - Math	Grade Batches for Each Item Spec Delivery		Archetypes		
		Number	Grade	Claim	Number	Grade
1	6	3, 7, HS		16	3,7,HS	1, 2, 3
2	15	3, 8, HS		18	3,8,HS	1, 2, 3, 4
3	14	7,8,HS		18	7,8,HS	1, 2
4	14	4,5,6		18	4,5,6	1, 2, 3, 4
5	13	4,8,HS		20	4,8,HS	1, 2, 3, 4
6	14	5,7,HS		18	5,7,HS	1, 2, 3, 4
7	12	5,6,HS		16	5,6,HS	1, 2, 3, 4
8	12	3,7,HS		18	3,7,HS	1, 2, 3, 4
9	13	5,6,HS		18	5,6,HS	1, 2, 3, 4
10	14	4,8		18	4,8	1, 2, 3, 4
11	12	3,HS		18	3,HS	1, 2, 3, 4
12	12	3,5,6		18	3,5,6	1, 2, 3, 4
13	10	4,7		19	4,7	1, 2, 3, 4
14	8	3,6		17	3,6	1, 2, 3, 4
	169			250		

Note: Archetypes were assembled to be representative of the entire set of items.

Note: The archetype numbers include any Performance Tasks that are developed as part of the archetype pool.

Table 4. English/Language Arts/Literacy Specifications and Archetype Delivery

ELA							
Delivery Number	Number of Specs - ELA	Grade Batches for Each Item Spec Delivery			Archetypes		
		Number	Grade	Claim	Target	Number	Claim
1	1	Stimulus	1-4				
1	7	3-HS	2	8			
2	7	3-HS	2	9			
3	28	3-HS	2-4	1, 2, 3			
4	28	3-HS	2, 4, PT	3, 4, 6, I/E PT			
5	55	3-HS	1, PT	1-7, N PT			
6	56	3-HS	1, PT	8-14, O/A PT			
7		3-HS			20	2	8, 9
8		3-HS			110	2-4	1-4, 6
9		3-HS			58	1	1-7
10		3-HS			62	1	8-14
Total	182	3-HS	Total		250		

Note: The archetype numbers include any Performance Tasks that were developed as part of the archetype pool.

Use of Systems and Tagging

The CTB Collaborative used DAS and ITS as the item authoring platforms to support production of various item types for the Smarter Balanced Assessment Consortium. Both systems were used for the authoring and review of stimuli and items/tasks for Smarter Balanced Contract 14. Clear procedures for the flow of items through authoring and review steps were developed and communicated to the various review groups and other stakeholders.

To support the implementation of the Smarter Balanced Field Test, the Collaborative ensured the availability of a robust set of item, task, and stimulus metadata to meet several purposes:

- Item tagging to support the computer-adaptive algorithm for test administration
- Content tagging to document full coverage of Smarter Balanced assessment claims and targets and the Common Core State Standards
- Item-bank tagging to support continued use of Smarter Balanced items by states
- Other item tagging for reporting and analysis

In addition to item-attribute tagging, items were associated with annotations to support Smarter Balanced accommodations. The expectation was that all metadata tags will be applied to items in the CTB DAS and the AIR ITS systems. This included annotations for

- Translations (including ASL)
- Braille
- Text-to-speech
- Glossaries (English and second-language)
- Other required accommodations tagging

The Collaborative was prepared to add/edit the list of item attribute tags that were under consideration for Smarter Balanced 16 item development and added tags as needed to meet the needs of the Field Test. This list, once approved, included additional tags to capture Smarter Balanced 16 requirements, such as Task Model, Specification Version Date, and Component Items for performance tasks. Other features required tagging (e.g. use of language complexity rubrics) were accommodated by additional tags.

Targeted training was provided to ensure that all item authors, developers, and reviewers understand the purpose and requirements for item attribute tagging.

Training Activities

Because of the large number of item and performance task writers that were ultimately involved in the development of the Field Test item pool, training involved live virtual sessions and the use of prerecorded modules that could be reviewed and accessed on demand.

In July 2013, the following activities were completed in preparation for training and professional development for educator item writers and editors:

1. Updates and revisions to training modules. While the modules developed for Contract 08 were effective in providing an introduction to the Smarter Balanced assessment system and components of item development, Contract 14 provided additional insight relative to aspects of the training modules that were most successful and what had to be updated and/or revised due to changes made throughout pilot test item and task development. For example, modules that focused on content and item specifications reflected general changes to those specifications as well as included sample items that reflect the current design and approach for Smarter Balanced items and tasks. Similarly, those modules that focused on item types were revised to address response types as indicated through item development conversations.

2. Development of new modules. The contractor collaborative developed new modules that focused on 1) the item authoring system(s), 2) in-depth training for writing items to each claim/target (including the use of item specifications, task models, and CCSS for each content area) and 3) expansion of accessibility considerations such as linguistic complexity. Those modules followed the design of those developed for Pilot Test item development.

3. Selection of educators for item writing. Recruitment began in early May 2013 to ensure parallel development across educator-, state-, and vendor-created items. Item writers were specifically recruited across a range of content areas, grade levels including higher education, and experience with under-represented student populations.

Virtual Training of Educator Item and Task Writers

Educator training sessions occurred after the school year had ended (mid-July), when participants had more time available. CTB experienced many scheduling conflicts and challenges during Contract 14 that they planned to circumvent via early, regular, and clear communication. For the virtual training workshops, video and/or audio presentations were consistent so that all educators hear the same messages from the Smarter Balanced work groups.

Continued Training during Development

Training and learning throughout item and task development was ongoing. Each educator worked directly with one of the CTB Collaborative assessment specialists. These specialists conducted regularly scheduled meetings in which educators could share and discuss challenges and successes. This approach would be similar to the model that was used during Contract 14. In this model, educators met and worked directly with one of the contractors' assessment specialists to receive appropriate support and guidance. For Contract 16/17, CTB was able to enhance that model through the added value of experienced educator item and task writers, who could also be called upon to provide feedback. Because item/task development was a phased process, retraining was conducted for the appropriate writers immediately prior to the onset of each phase. This was "just-in-time" training that focused on the content and item specifications. If issues arose during item development, The Collaborative conducted ad hoc training for individuals, small groups, or large groups, as needed.

Certification and Management

At the onset of item development, all item development entities (subcontractors, external item development vendors, states, and individual educators) were required to develop the appropriate samples needed for item development certification. CTB provided preliminary training for individual educators recruited as item writers prior to assigning the 20-item certification set. Once potential item developers met the preliminary qualifications, each organization or individual wishing to continue in the certification process created a sample set of 20 items reflective of the anticipated item assignments. This requirement applied to item and task development organizations within our collaborative (CTB, AIR, DRC, MI, and SCALE), any external item and task vendors used for item authoring, groups of educator authors within a state, and individual educator item authors who were recruited directly.

The table below outlines a sample certification set for item authors or organizations seeking approval to move forward with item development. In this example, the organization would create six 20-item sets for approval prior to the start of item development for Claims 1 and 4.

Table 5. Sample Certification Set

Claim	Grade Band		
	3-5	6-8	HS
1	X	X	X
2			
3			
4	X	X	X

Item/Task Set Evaluation

Once the sample item or task set was created, senior content reviewers evaluated the item set based on the approved criteria for certification. The contractor provided initial feedback or additional training to item developers, as needed. Once an item set met all criteria, the item developer was recommended for certification.

All item development entities (subcontractors, external item development vendors, states, and individual educators) were required to develop the appropriate samples needed for item development certification. This 20-item certification set was reflective of the anticipated item assignments. This set of items was reviewed by the collaborative senior reviewers, as well as by Smarter Balanced representatives (SMEs, Accessibility, Sensitivity and Bias experts) for adherence to the item quality review criteria for Smarter item development. For an entity to qualify for Smarter item development, 95% of the items in the certification set and in all sets of items submitted for the Smarter 16 Field Test pool had to meet an acceptance of 95%.

Role of Item Quality Review Panel

The Item Quality Review Panel (IQRP) was recruited in early May 2013 and consisted of seven panelists for ELA and nine members for math. Panel members are content experts or those who inform decisions related to students with disabilities and English Language Learners. The panel gave feedback during reviews of specifications, archetypes, and item and performance task batches.

The IQRP held an initial Face-to-face meeting in late May 2013. The outcomes of this meeting included recommendations to consider for item development and contributions to the item quality-review criteria. The recommendations from this meeting were refined and implemented based on Smarter Balanced confirmation and agreement to the overall vision of the assessment.

Educator Recruitment Activities

Smarter Balanced recruited educators for each of the assessment activities listed in Table 6 below. The inclusion of qualified educators in these assessment activities builds capacity and creates sustainability for the Smarter Balanced assessment system. The contractor, with assistance from the governing states' Teacher Involvement Coordinators (TICs), will sought an educator sample that ensured a balanced representation based on grade levels, content area, and other demographic data across the governing states. For each recruitment activity, TICs used an information packet describing the purpose of the activity and recruitment process. The information packets included the following:

- Documentation for TICs about the recruitment activity, specifically state recruitment and qualifications for educators
- A summary of overall recruitment processes in the coming months
- Specific counts of educators CTB is recruiting from each state for the educator opportunity
- A list of Frequently Asked Questions that may be used by TICs to provide information about educator opportunities
- A sample educator recruitment email that TICs may use in support of state recruitment. State-specific information can be added or clarified in this template.

In addition to requests for specific numbers of educators across content areas and grades, TICs received specific targets for the recruitment of educators with experience in working with under-represented student populations (English language learners and students with disabilities).

The recruitment of educators for each of the activities occurs in phases. The recruiting activities and timelines are provided in Table 6.

Table 6. Recruiting Activities and Timeline

Activity	Number of Educators Required	Information Package(s) to TIC	Timeframe for Recruiting	Notification to TICs of Participants and Alternates	Timeframe for Activity
Phase I					
Pilot Test Range Finding	105	5/13/13	5/14 to 5/28/13	5/30/13	June
Field Test Item and Task Authoring	200	5/15/13	5/16 to 5/30/13	6/4/13	June–October
Phase II					
Pilot Test Item-Data Review	28	5/22/13	5/23 to 6/6/13	6/11/13	August–September
Field Test Stimulus Review	284				June–August
Field Test Item-and-Task Review	296				July–November
Phase III					
Alignment Study	TBD	11/4/13	11/5 to 11/18/13	12/11/13	January–March
Phase IV					
Field Test Range Finding	TBD	2014	2014	2014	TBD

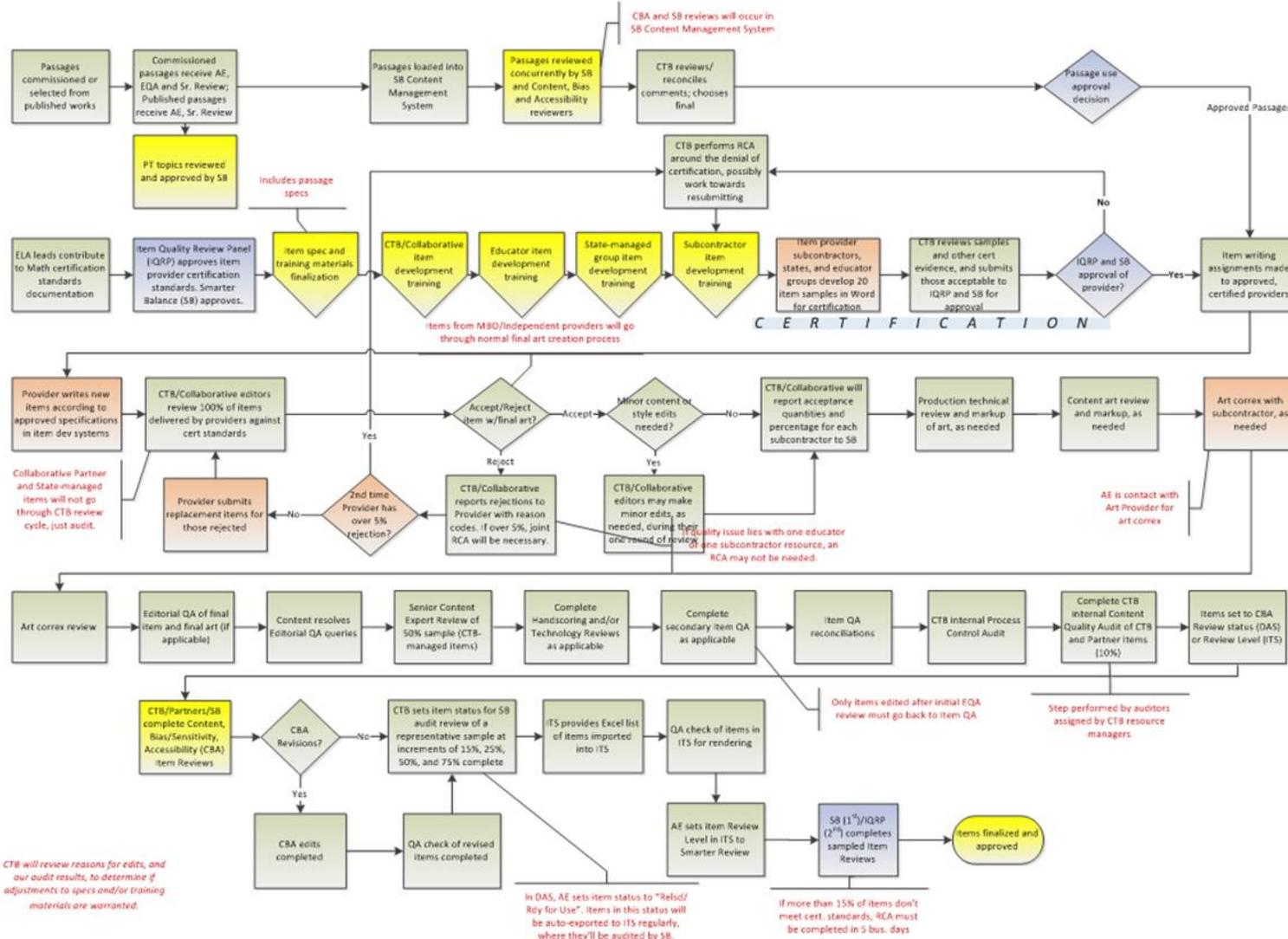
Item development process

The charts below outline the detailed process for stages of item development. They describe the many checks and reviews each item receives before it is approved for field testing. Item content, graphics, artwork, response processes and stimuli get extensive reviews. Items are also subject to reviews for possible cultural bias or material that may distract some test takers because it is in an area of sensitivity. Throughout the process there are checks to assure that items are accessible to as many students as possible.

ELA ITEM DEVELOPMENT PROCESS – SMARTER BALANCED 16

Content Provider	CTB/ Collaborative	Smarter Balanced	Joint Tasks
------------------	--------------------	------------------	-------------

Assumption: All ELA item development completed in DAS or ITS development system (SB Content Management System)



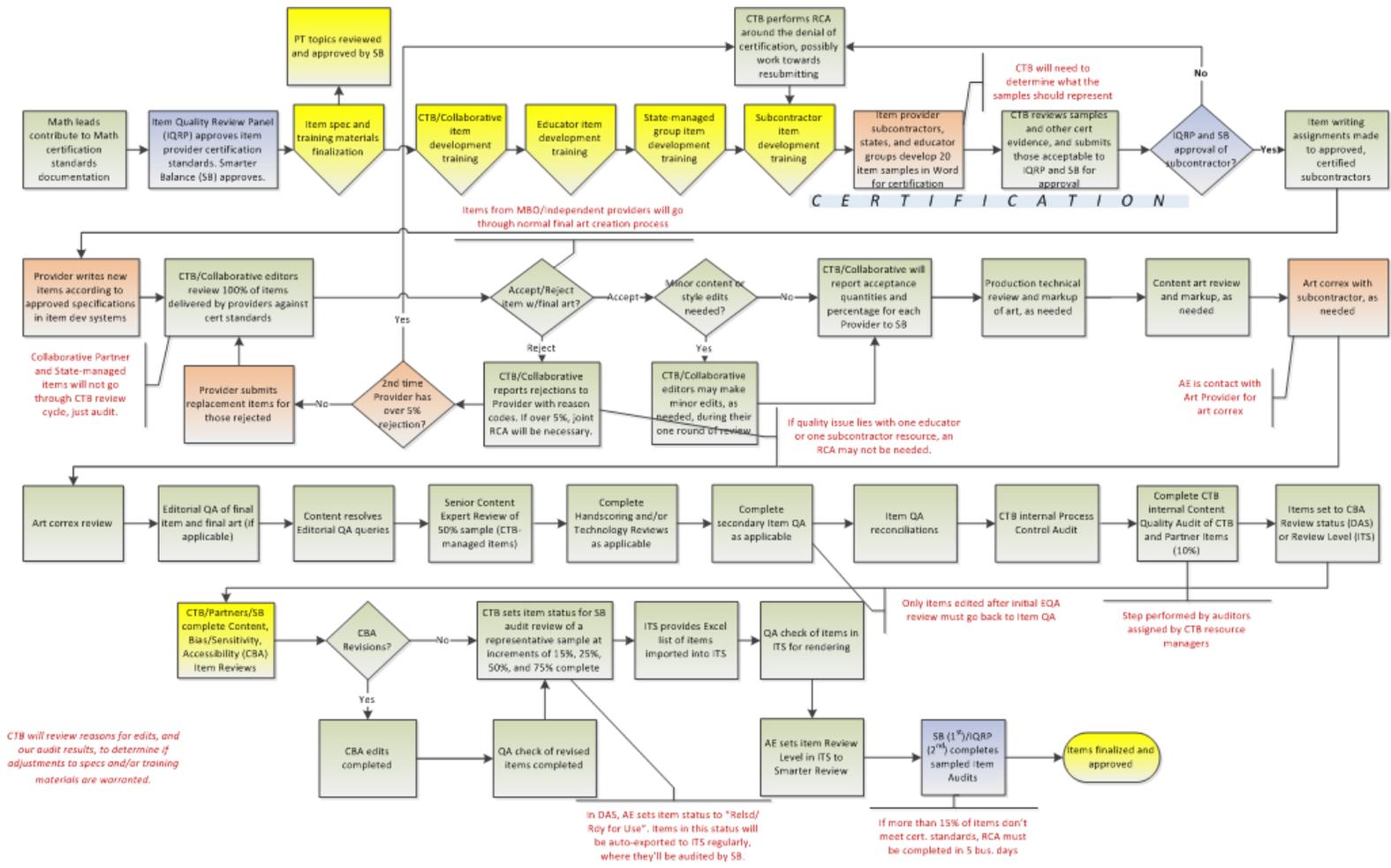
CTB will review reasons for edits, and our audit results, to determine if adjustments to specs and/or training materials are warranted.

In DAS, AE sets item status to "Rebid/Roy for Use". Items in this status will be auto-exported to ITS regularly, where they'll be audited by SB.

If more than 15% of items don't meet cert. standards, RCA must be completed in 5 bus. days

Assumption: All MATH item development completed in DAS or ITS item development system (SB Content Management System)

MATH ITEM DEVELOPMENT PROCESS – SMARTER BALANCED 16



Detailed information about item writing, development, review and scoring can be obtained upon request.

Table 7. Additional item writing, development, review and scoring documentation

Topic	Sub-topic	Document Name
Item Writing	Process Flow	20150512 Item Development Process Description FINAL
		20150512 Smarter process maps FINAL
		Smarter 16 ITS Final Content Approval checklist FINAL
		Smarter 16 Final Web Approval Checklist20150512
	Models-Specifications	20131003 Smarter 16 Item pool specification v12a Math FINALnew
		20131006 Smarter 16 Item pool specification v12d ELA FINALnew
		ELA Archetypes Math_Archetype_Metadata
	Review criteria	SB_16_ELA_Quality_Criteria_FINAL
		SB_16_MATH_Quality_Criteria_FINAL
		CBA Item Review Business Rules 9-25
Human Scoring	Process Description	20150512 Smarter Hand Scoring Process FINAL
	Qualifications	20150512 Smarter Hand Scoring Rater Qualifications FINAL
	Quality Monitoring	20150512 Smarter Hand Scoring Quality Monitoring FINAL
	Recruitment-Training	0150512 Smarter Hand Scoring Rater Training FINAL
	Data Review	20150512 Smarter 2014 Field Test Data Review Summary Report FINAL
		20150512 Smarter Data Review Results Summary