Hand-scoring Rules

Overview

In order to optimize scoring reliability, scoring validity and efficiency, Smarter Balanced has provided this scoring method recommendation, but the scoring process will ultimately be a state determination. Smarter Balanced provides expectations for reliability and validity that member states must meet or exceed. Smarter Balanced intends to use reliability and validity criteria that are consistent with industry best practices as established as a deliverable of Smarter Balanced Contract-16/17. Smarter Balanced has posted question and answer documentation to SmarterApp.org regarding Smarter Balanced Assessments.

In order to control costs associated with hand-scoring, Smarter Balanced has reimagined the training process for item raters. Instead of training raters for every individual item, Smarter Balanced will train raters on families of items called Task Models. In ELA, qualification will still occur at the item level. In mathematics, raters will qualify at the task level for most items. Smarter Balanced expects vendors to maintain traditional item-level hand-scoring statistics, including rater agreement and inter-rate reliability rates.

ELA. In order to score ELA items, raters will receive training at the level of the task model. Smarter Balanced will provide "baseline" anchor and training sets as well as rubrics by writing purpose (e.g., informative writing) for full-write essay items. Qualification and validation sets will be provided for each essay. Anchor and training sets will also be provided for the task models associated with the ELA short text items in the CAT and PT sections. For the ELA short text items in the CAT and the PT sections, raters will receive training by grade span (grades 3-5, 6-8, and high school) instead of by grade level. Smarter Balanced estimates the number of rater trainings as follows:

- ELA Essay: 7 trainings by grade level, grades 3 – 8
- ELA Essay: 5 trainings for high school
- ELA short text items: 16 trainings by grade span (3-5, 6-8, and high school)

Even though training is at the task-model level, qualification will occur on an item-by-item basis for all ELA handscored item types, and Smarter Balanced will provide qualification and validation sets for each item in ELA. For those item types where raters were trained by grade span, raters will qualify on each item within a specific grade.

Mathematics. In order to score mathematics items, raters will receive training and will qualify on task models for almost all items. For Mathematics, Smarter Balanced will provide anchors and training sets for the task models. Smarter Balanced will provide item-specific rubrics and item-specific validation sets for all Mathematics items. Smarter Balanced estimates that there will be between 15 and 20 rater trainings for the mathematics PT items by grade level. Approximately 82 items will need individual training and qualification across all grades within mathematics.

Hand-Scoring Rules

Security

Site Policies

Smarter Balanced requires sophisticated systems for securing examination materials and other customer data. Security provisions must be implemented to ensure the security of all assessment materials, including examination questions, student responses, and training material. These materials are considered highly secure, and appropriate policies should be implemented to govern access, transfer (hard-copy and digital), and storage and archival

Facilities are to be secure and access to buildings is restricted to authorized personnel only. All personnel are required to sign the applicable security agreement and sign confidentiality/non-disclosure forms. Each scoring contractor must ensure that all nondisclosures have been signed and that a copy is maintained on file. Violation of any security procedure or policy or non-disclosure agreements must be reported immediately by the scoring contractor to Smarter Balanced.

Restrictions Regarding Communication

All information regarding Smarter Balanced Field Test is to be treated with the highest security. Confidentiality/non-disclosure agreements restrict all external communication regarding test items. All employees are directed to have no contact with the press or other media regarding Smarter Balanced or other hand-scoring procedures. In addition, readers are informed they are not allowed to discuss Smarter Balanced hand-scoring information outside of the specified training/scoring areas or outside of the building. Re-creating any portion of the test or responses is also prohibited.

All material, including training material, is to be considered secure, unless otherwise directed by Smarter Balanced. Items, stimuli, student responses, or any student information must not be communicated via e-mail. The preferred mode of communicating is delivery via the FTP (File Transfer Protocol) site. Hand-scoring personnel will follow the guidelines outlined in the most current version of “Contractor Orientation: A Presentation and Reference Guide for Smarter Balanced Contractors.”

Hand-Scoring Rules

Scoring Design

Approximately 1,869 items were scored during the Smarter Balanced spring 2014 Field Test. The table below will be updated after the completion of Data Review to provide estimated counts for the Summative, Interim Comprehensive, and Interim Assessment Blocks.

Table 1: Hand-Scoring Estimated Item Count Summary (as of 3/14/14)

	CAT Items			Performance Tasks		
Grade	**ELA Reading Short-Text Items**	**ELA Brief Writes**	**Math Short-Text Items**	**ELA Research Short-Text Items**	**ELA Full Writes**	**Math Short-Text Fill-In Table Items**
Grade 3	43	28	0	38	19	86
Grade 4	40	27	6	48	24	94
Grade 5	45	31	6	50	25	77
Grade 6	43	30	3	39	19	86
Grade 7	41	27	25	50	25	76
Grade 8	49	29	16	54	27	66
High School	157	97	43	56	28	96
Totals	**418**	**269**	**99**	**335**	**167**	**581**

Training Process

Training Process: ELA/Literacy Performance Task Full Writes

Baseline anchor sets, from the Smarter Balanced Pilot full-write tasks, were developed for each grade and writing purpose. The anchor sets will be used to train readers on each of the writing traits— Conventions, Organization/Purpose, and Evidence/Elaboration or Development/ Elaboration— at a particular grade level. For all writing purposes, Organization/Purpose is the first trait and Conventions is the third trait. Evidence/Elaboration is the second trait for the opinion, argumentative, informational, and explanatory writing purposes. Development/Elaboration is the second trait for the narrative writing purpose. The number of full-write tasks at each grade and purpose are shown in Table 2.

Table 2: Field Test Full Write Summary by Writing Purpose (Standard Setting and Census)

Grade	Informational (Gr 3 5) Explanatory (Gr 6 HS)	Opinion (Gr 3 5) Argumentative (Gr 6 HS)	Narrative	Total
3	8	4	7	19
4	9	5	10	24
5	10	4	11	25
6	12	3	4	19
7	9	11	5	25
8	9	12	6	27
High School	15	13	N/A	28
Totals	**72**	**52**	**43**	**167**

Appendix A diagrams the training process for ELA/literacy full writes.

Below is a description of the training steps.

1. Trainer introduces ELA/literacy performance tasks, their general structure, purposes and full-write rubrics.
2. Trainer presents the Conventions rubric for the grade level as well as the appropriate Conventions charts (for all grades up to and including the grade being trained).
3. Trainer explains the scoring rationale for each of the anchors in the Conventions baseline anchor set (Steps 1-3 are performed only once for each grade level).

4. Trainees read the task and source materials the first Field Test full write in the grade level and purpose category (for example, Grade 3 Informational).
5. Trainees score the training set (5 papers) for Conventions, Full Write 1.
6. Trainer reviews the correct scores and scoring rationale for the Conventions training set.
7. Trainer presents the Organization/Purpose rubric in relation to the writing purpose (Narrative, Informational, Explanatory, Opinion, or Argumentative).
8. Trainer explains the scoring rationale for each of anchors in the Organization/Purpose baseline anchor set.
9. Trainees score the training set (5 papers) for Organization/Purpose, Full Write 1.
10. Trainer reviews the correct scores and scoring rationale for the Organization/Purpose training set.
11. Trainer presents the Evidence/Elaboration rubric or Development/Elaboration rubric, and Elaboration Guidelines specific to the writing purpose.
12. Trainer explains the score rationale for each of the anchors in the Evidence/Elaboration or Development/Elaboration baseline anchor set emphasizing the difference between Organization/Purpose and Evidence/Elaboration.
13. Trainees score the training set (5 papers) for Evidence/Elaboration or Development/Elaboration, Full Write 1.
14. Trainer reviews the correct scores and scoring rationale for the Evidence/Elaboration or Development/Elaboration training set.
15. Trainer checks for understanding and reiterates the distinctions between the two traits.
16. Trainees score a fourth training set (5 papers) for all traits combined (i.e., readers score traits 1, 2, and 3 for paper 1, then traits 1, 2, and 3 for paper 2, etc.).
17. Trainer explains the scoring rationale for the training responses and answers questions
18. Trainees score a qualification round (10 papers) for all three traits for Full Write 1.
19. Trainees who do not meet the qualification standard on round 1, score round 2 (See Table 4)
20. Qualified readers begin scoring.

Material for full-write training, at each grade level, will include:

1. Baseline Anchor Sets approved during Smarter Balanced Pre-Range-Finding
2. Field Test Tasks and Source Materials
3. Purpose/Task Specific Rubrics
4. Conventions Charts (Smarter Balanced approved)
5. Supplemental Scoring Guidelines (Smarter Balanced approved)
6. Training sets (specific to the first full-write task for each grade/purpose)
7. Qualification Rounds (generally administered in 2 rounds of approximately 10 responses per full-write task)

Training Process: ELA/Literacy Short-Text Items

Readers will be trained by grade band (3-5, 6-8, or High School) and by claim and target subcategory as described in Table 3.

Table 3: Field Test ELA/literacy Short-Text Item Types and Subcategories

Item Type	Subcategories
Claim 1 Reading (CAT Short Text)	Training will occur at the grade-band level for the following subcategories: 1. Target 2 (Central Idea, Literary) 2. Target 9 (Central Idea, Informational) 3. Target 4 (Inference/evidence, Literary) 4. Target 11 (Inference/evidence, Informational)
Claim 2 Brief Writes (CAT Short Text)	Training will occur at the grade-band level for the following subcategories: 1. Target 1a- Narrative a. Organization, Opening b. Organization, Conclusion c. Elaboration 2. Target 3a- Informational-Explanatory a. Organization, Introduction b. Organization, Conclusion c. Elaboration 3. Target 6a- Opinion-Argumentative a. Organization, Introduction b. Organization, Conclusion c. Elaboration
Claim 4 Research (Performance Task Short Text)	Training will occur at the grade-band level for the following subcategories: 1. Target 2 (Interpret and Integrate Information) 2. Target 3 (Evaluate Information/Sources) 3. Target 4 (Use Evidence)

Appendices B, C and D diagram the training process for ELA/literacy short-text items.

Below is a description of the training steps.

1. Trainer introduces the short-text category— the claim, target, and additional subcategory (for Claim 2).

2. Trainer presents the short-text items that are represented in the anchor and training sets, any associated source materials or stimuli, and the item-specific rubrics.
3. Trainees read the associated source materials or stimuli as appropriate.
4. Trainer explains the scoring rationale for each of the anchors (anchor set for the claim/target/subcategory).
5. Trainees score the training set (5-10 papers) for the short-text claim/target/subcategory.
6. Trainer reviews the correct scores and scoring rationale for the training set.
7. Trainer checks for understanding and answers questions.
8. Trainees read the prompt, source materials, or stimuli for the first short-text item in the claim/target/subcategory (example Grade 6, Claim 1 Reading Item 1).
9. Trainees score a qualification round.
10. Qualified readers begin scoring.

Material for short-text item training will include:

1. Anchors and Training Sets (by grade-band/claim/target/subcategory)
2. Prompts and Source Materials or Stimuli
3. Item-Specific Rubrics
4. One Qualification Round (10 responses per item)

Training Process: Mathematics Items Appendices E and F diagram the training process for Mathematics.

Below is a description of the training steps.

1. Trainer introduces the performance tasks and their structure, or the CAT short-text items.
2. Trainer presents the items that are represented in the anchor and training sets, any associated source materials or stimuli, and the item-specific rubrics.
3. Trainees read the associated source materials or stimuli, as appropriate.
4. Trainer explains the scoring rationale for each of the anchors (anchor set for the performance task grade/family/item category or the CAT item category).
5. Trainees score the training set for the item category, as described in step 4.
6. Trainer reviews the correct scores and scoring rationale for the training set.
7. Trainer checks for understanding and answers questions.
8. Trainees score a qualification round.
9. Qualified readers begin scoring.

Material for Mathematics training will include:

1. Anchors and Training Sets (by performance task grade/family/item category or by CAT item)
2. Prompts and Source Materials or Stimuli
3. Item Specific Rubrics
4. One or two Qualification Rounds per item category, depending on item complexity (10 responses per round,)

A further note on Mathematics Performance Tasks:
Unlike ELA Performance Tasks, Mathematics Performance Tasks (PTs) contain interdependencies among the items in a task. Each Mathematics PT is made up of six items. The first two items are generally machine-scored items. Of the remaining four items, two to four of the items are hand-scored. (Independent machine-scorable items tend not to be hand-scored.) Items may be dependent on any of the previous items in the PT.

For example, item #6 may build on items #3 and #5. The rubric for item #6 will specify the correct response based on prior correct responses to items #3 and #5. Readers are responsible for determining the appropriate response to item #6, and awarding credit accordingly, when the student's responses to items #3 and #5 are incorrect.

The proper handling of tasks with dependencies must be addressed in training and readers should have practice with working through a student's PT responses and recognizing correct work based on previous incorrect values. In general, training materials are organized so that readers train on a task model rather than on a complete PT. When these items are presented in training though, any item that may be a dependency (even if it belongs to a different task model) is included for scoring reference.

Qualification Process

After the training set has been completed for a full-write task or an item category, the reader qualification process begins (see Reader Training Models in Appendices A-E). All team leaders and readers will be required to qualify before scoring and will be informed of what they are expected to achieve in order to qualify. The standards, provided in Table 4, are qualification expectations for these score point ranges and response types. A reader is required to meet the qualification standards on one qualification round in order to score Field Test student responses.

Qualification should be conducted through an online system so that the results can be electronically captured for each individual trainee and their scoring assignments can be reliably made.

Table 4. Field Test Qualification Standards

Score Point Range	Qualification Standard (Exact Agreement)
0-1	90%; no non-adjacent scores
0-2	80%; no non-adjacent scores
0-3	80%; no non-adjacent scores
0-4	70%, no non-adjacent scores

Condition Codes

Because condition-code responses are systematically routed to scoring supervisors for final code assignment, supervisors require detailed training on the Smarter Balanced condition codes and definitions.

The following condition codes will be valid used for scoring. Condition code categories are also listed in Appendix G.

Table 5. Field Test Condition Codes

Condition Code	Condition Code Category
B	**Blank**
I	**Insufficient** (a) Student has not provided a meaningful response Some examples: • Random keystrokes • Undecipherable text • *I hate this test* • *I don't know, IDK* • *I don't care,* • *I like pizza!* (in response to a reading passage about helicopters) • Response consists entirely of profanity (b) For ELA Full Writes, use "I" code (Insufficient) for responses described above and also if: Student's original work is insufficient to make a determination whether he or she is able to organize, cite evidence/elaborate, and use conventions as defined in the rubrics, or Response is too brief to make a determination regarding whether it is on purpose or on topic
L	**Non-Scorable Language** ELA/literacy: Language other than English Mathematics: Language other than English or Spanish
T	**Off Topic** *for ELA Full Writes Only* Definition: A writing sample will be judged "off topic" when the response is unrelated to the task or the sources or shows no evidence that the student has read the task or the sources (especially for informational/explanatory and opinion/argumentative) • "Off topic" responses are generally substantial responses
M	**Off Purpose** *for ELA Full Writes Only* Definition: A writing sample will be judged off purpose when the student has clearly not written to the purpose designated in the task. • An off-purpose response addresses the topic of the task but not the purpose of the task • Note that students may use narrative techniques in an explanatory essay or use argumentative/persuasive techniques to explain, for example, and still be on

	purpose • Off purpose responses are generally developed responses (essays, poems, etc.) clearly not written to the designated purpose

Alerts

Readers are also trained to watch for indications of "troubled students" and/or cheating. Such information can require urgent attention prior to the completion of hand-scoring. A student response to any hand-scored test item that may be of a sensitive nature should be assigned a score and identified as an "Alert." The different types of alerts are listed below.

Troubled Student Alerts.

Troubled Student Alerts include, but are not limited to, the following:

- Suicide
- Criminal activity
- Alcohol or drug use
- Extreme depression
- Violence
- Rape, sexual, or physical abuse
- Self-harm or intent to harm others
- Neglect

Testing Irregularities.

Testing Irregularities include, but are not limited to, the following:

- Assistance/Intervention: The student admits to having received assistance during the test administration.

The communication for Alert papers should adhere to State policy, but should follow the following tenets to ensure immediate notification is completed:

- For Troubled Student Alerts, the State Testing Director for the flagged student will be notified the same day via a phone call, and a copy of the response and the student ID will be sent via overnight mail. (See Appendix H for the sample notification letter.) Responses indicating "testing irregularities" will be logged and sent to the State Testing Director and Smarter Balanced at the conclusion of scoring.

Scoring Process

The scoring model will be based on scoring one item at a time (i.e. in single-item Rater Item Blocks). For some Mathematics performance task items, where there are scoring dependencies on previous items, Hand-Scoring may determine that it is more efficient to score those items together in a Rater Item Block (RIB). Additionally, the three-trait full write tasks will be scored in a single RIB by one reader.

Each item, or RIB, will be scored by a (recommended) team of five to ten readers under the direction of a team leader and a supervisor. Each supervisor will be responsible for multiple teams in a specific content and grade band. These teams will also be monitored by a team leader and

supervisor. Hand-Scoring staff will ensure that, at any given time, no fewer than three readers are scoring an item. Each reader will work at his or her own computer station to read student response and enter a score for each item.

For monitoring inter-rater reliability, a 10% second read rate will be required. When an item response requires a second read, the response will be placed back into the scoring queue for a second blind score. The second reader will be unaware of the first reader's score. The first and second reads should be randomly distributed by the scoring applications.

Quality Monitoring

The types of data used and the processes applied for ensuring accuracy to maintain standards for rater quality and item data quality to should be consistent across Smarter Balanced states. Processes will include administration of validity sets, monitoring validity and inter-rater reliability data, and resetting scores when anomalous scoring patterns are identified.

A diagram of the scoring and quality monitoring processes is in Appendix I.

Validity Papers

One of the key measures of reader accuracy is agreement rates on validity papers. Validity papers assist in monitoring scoring accuracy and maintain a consistent focus on the established rubric and guidelines. These papers are pre-scored and pre-approved by hand-scoring supervisors or designated team leaders and are administered to team leaders and readers during each shift. Supervisors will ensure that validity papers are submitted appropriately so that all readers score the available validity sets, regardless of the individual reader's scoring pace.

Validity papers will appear to readers and team leaders in the same format as do actual student responses. The scores assigned to the validity papers are compared to the approved score, and information is obtained through this comparison about the accuracy and reliability of the reader and whether the scoring team/individuals are drifting from the original score criteria. Hand-Scoring managers, or a designated quality monitor, will check the score frequencies in each validity set pool to ensure that the possible range of score points is represented by the validity responses for each item.

Review of incorrectly scored validity papers is an example of a corrective action that may be prescribed by the data monitor based on analysis of validity and/or inter-rater reliability data. This review is an ongoing process that is conducted to maintain adherence to scoring criteria throughout the scoring effort. Routine procedures may include room-wide discussions led by the hand-scoring supervisor and team discussions conducted by team leaders, as well as one-on-one sessions with individual readers.

Readers who do not maintain the pre-determined scoring quality standard are retrained or disqualified from scoring the item or task in question. The quality standards are shown in Table 6 below and are based on the qualification standards the readers must meet in order to score (Table 4). When a reader's performance falls below the validity quality standard, the data monitor and/or scoring manager will review the reader's scoring patterns for the validity data and inter-rater reliability data from the time period in question and make a determination about rescoring the reader's work.

Table 6. Exact Agreement Standards

Score Point Range	Validity Standard (Exact Agreement)
0-1	90%
0-2	80%
0-3	80%
0-4	70%

Inter-rater Reliability

Inter-rater reliability reports indicate agreement rates between readers and provide a picture of readers' scoring patterns. Analysis of the inter-rater reliability reports is an excellent source to determine team or reader drift and team leader influence. The supervisor can re-anchor readers or teams with live papers, relevant training materials, and scoring guidelines.

Hand-Scoring Reports

Inter-rater reliability reports and validity reports will be used daily to monitor reader performance and provide feedback to readers.

Supervisors will review hand-scoring reports each day and will develop strategies to correct any problems that are revealed by the reports.

Quality data includes:

- Inter-rater reliability (item level and reader level with exact, adjacent, and non-adjacent rates)
- Validity or checkset results (item level and reader level with exact, adjacent, and non-adjacent rates)
- Item-level and reader-level reports on item score-point frequencies, including non-scorable codes
- Item-level reports showing mean scores

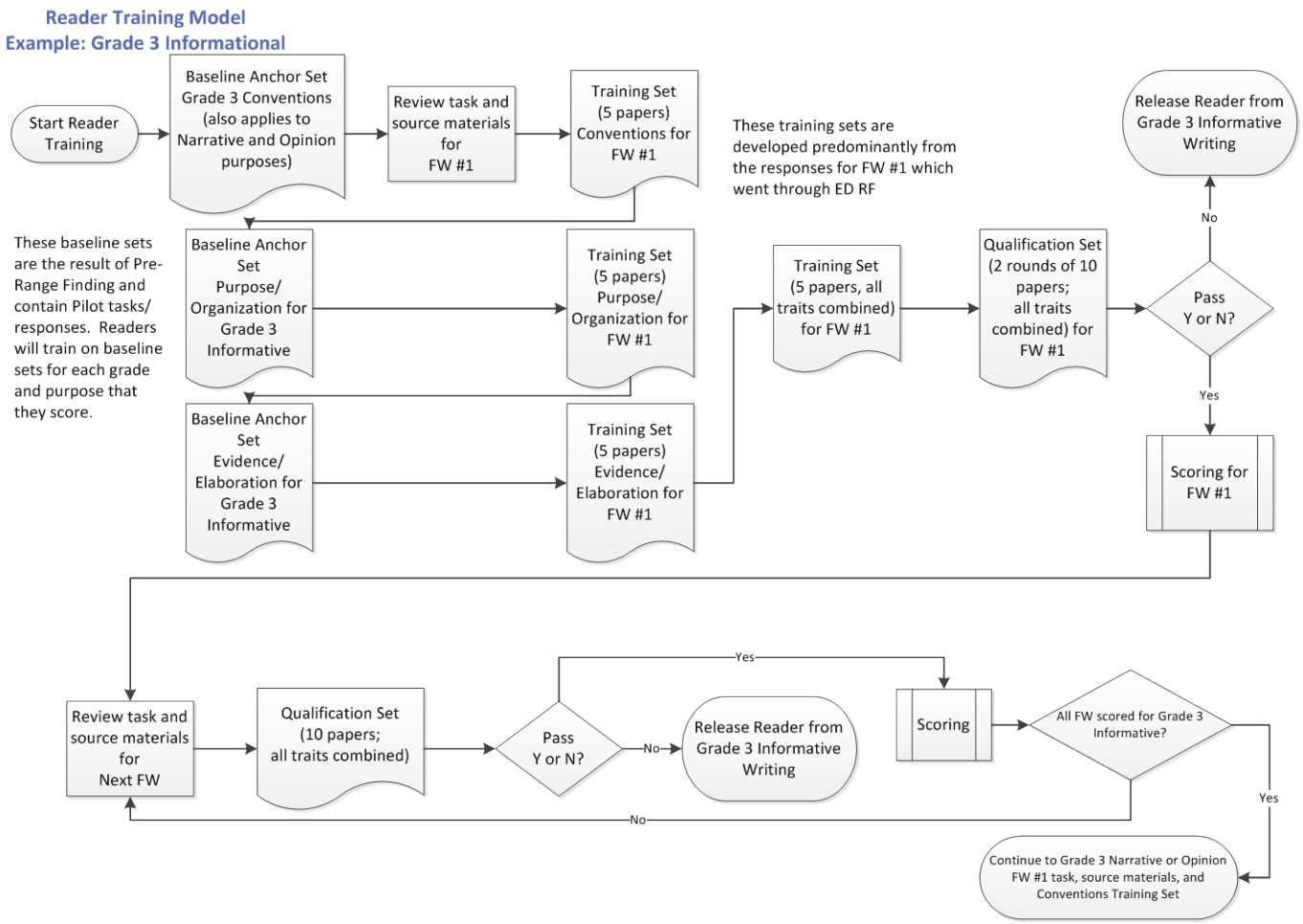
Recommendations

The following are recommendations for the Operational administration of the Smarter Balanced hand-scored items.

1. Augment the Validity sets to maintain no less than 30 responses per operational item.
    a. It is recommended that, for operational scoring, readers score a minimum of 10 unique validity sets per day per item so that there is sufficient performance data to recognize errant scoring trends
    b. Validity sets should include an occasional condition-coded response
2. Require readers to pass a qualification round on condition codes prior to any operational scoring; this is probably most critical for ELA full writes where readers need clear direction on the distinction between non-scorable responses and condition code responses

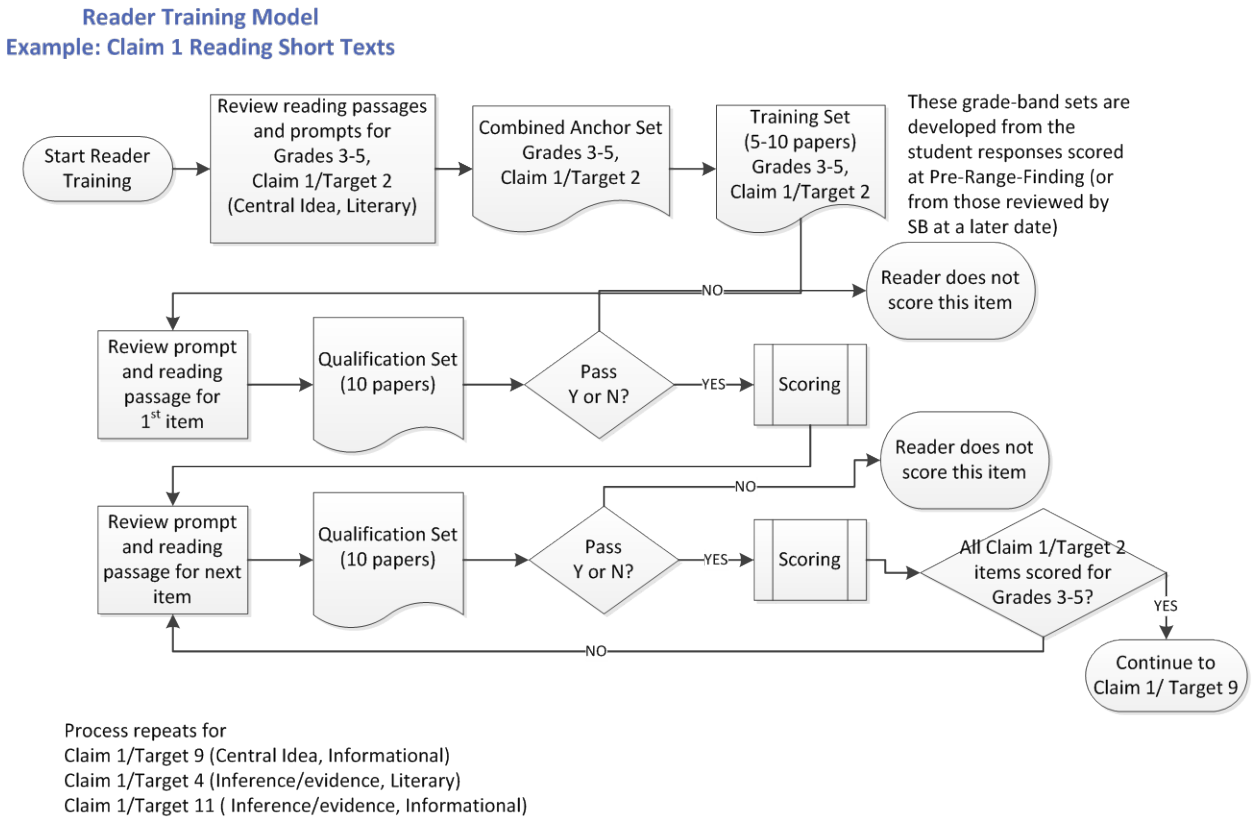
3. Consider augmenting item rubrics with anchor responses (for items that were not "baseline" items in the Field Test) (Note: This recommendation is under review by Smarter Balanced at this time)
4. For quality monitoring purposes, consider including additional data and/or reports
    a. Reader-level high and low adjacencies/discrepancies in inter-rater reliability
    b. Reader-level high and low adjacencies/discrepancies in validity set scores
    c. Detailed validity set reports that indicate, at a reader level, which validity set score point(s) are commonly missed

Appendix A: ELA/Literacy Performance Task Full Writes, Reader Training Model



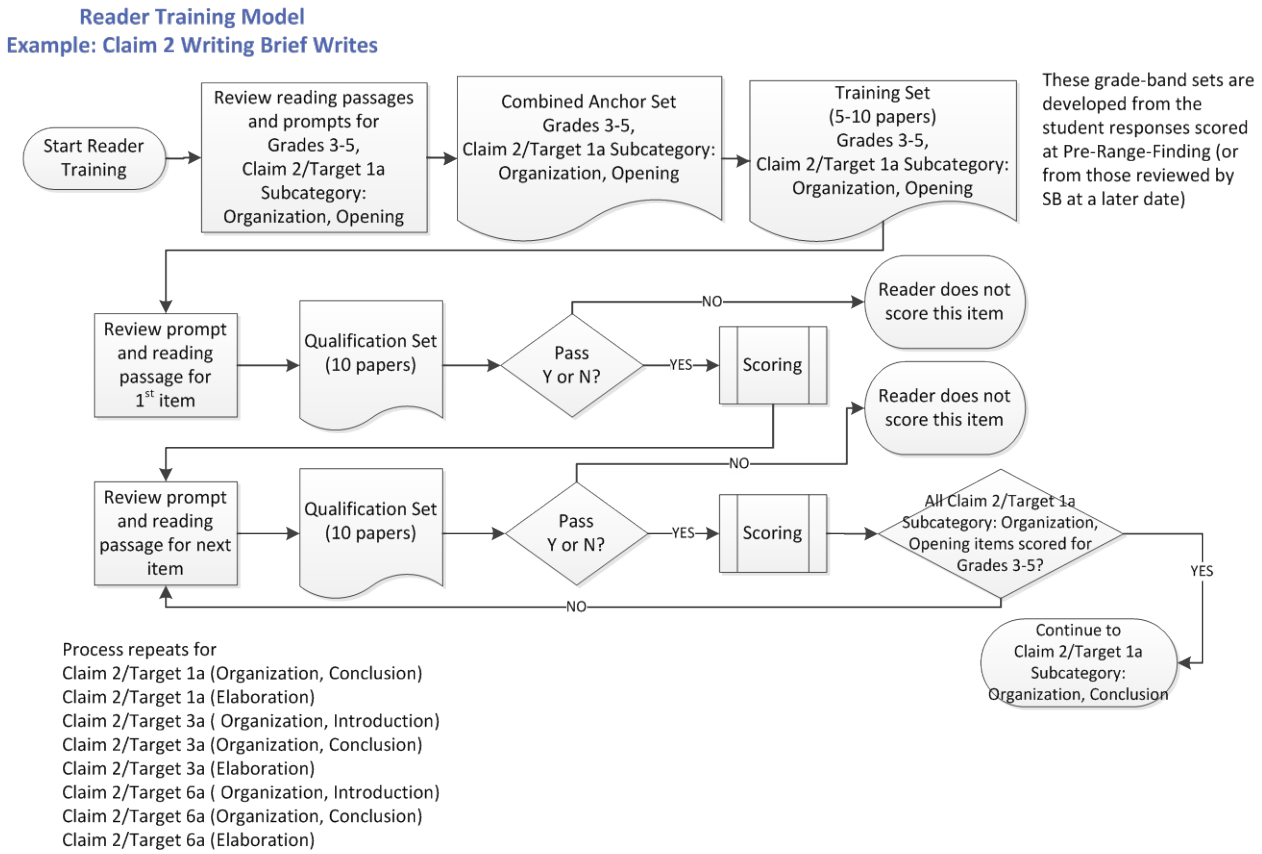
"Pass Y or N?" refers to whether or not the reader meets the qualification standard for the item's score point range.

Appendix B: ELA/Literacy Short-Text Items, Claim 1 Reading, Reader Training Model



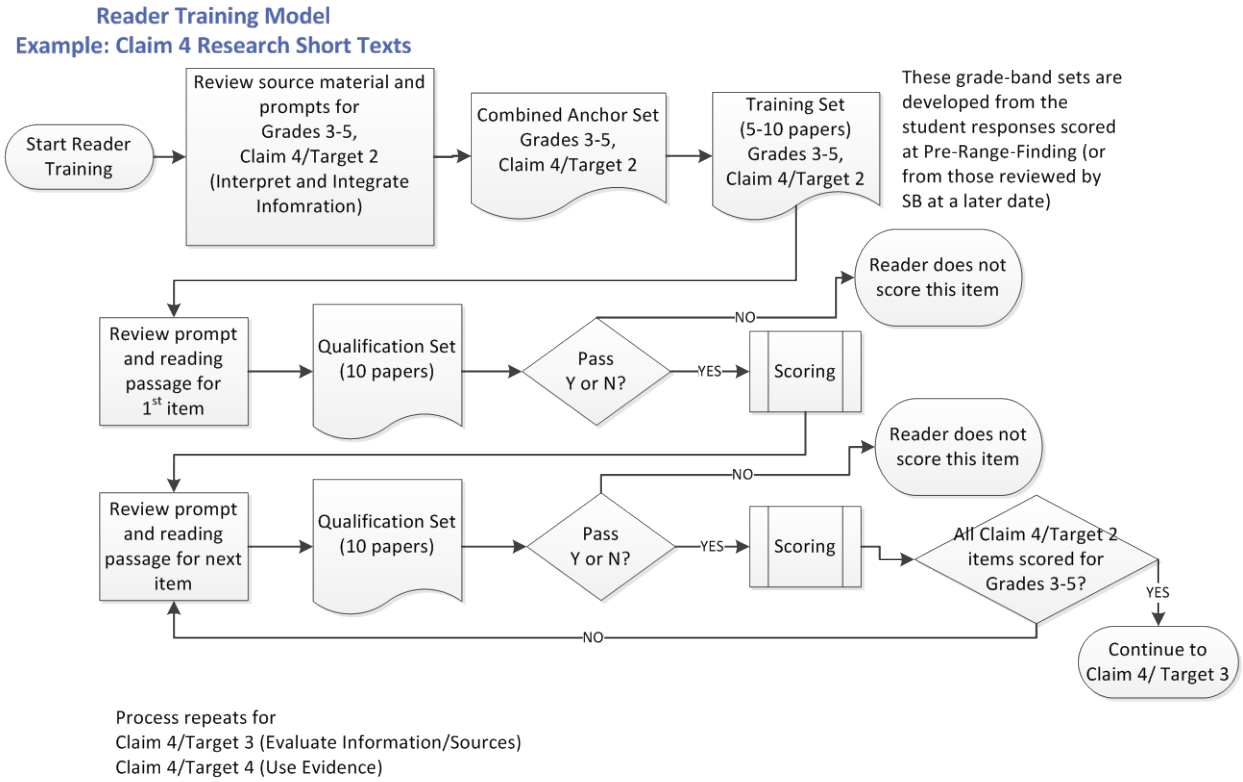
"Pass Y or N?" refers to whether or not the reader meets the qualification standard for the item's score point range.

Appendix C: ELA/Literacy Short-Text Items, Claim 2 Brief Writes, Reader Training Model



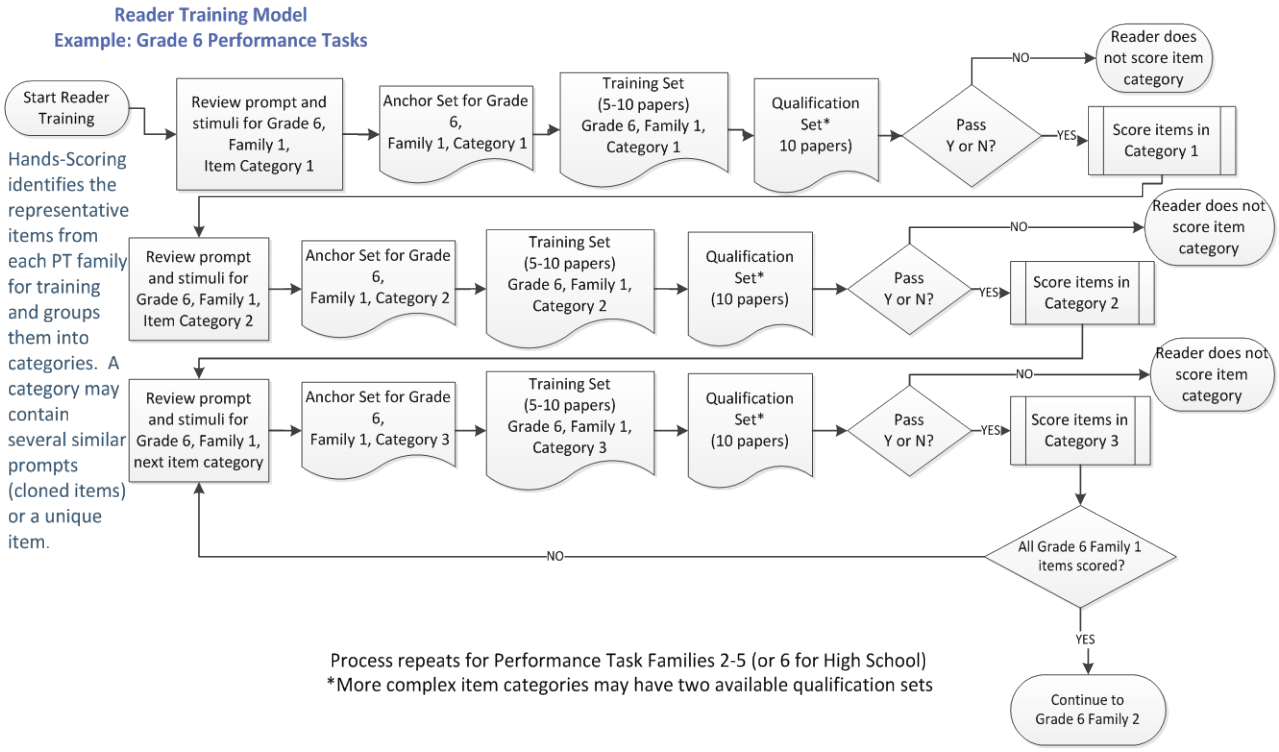
"Pass Y or N?" refers to whether or not the reader meets the qualification standard for the item's score point range.

Appendix D: ELA/Literacy Short-Text Items, Claim 4 Research, Reader Training Model



"Pass Y or N?" refers to whether or not the reader meets the qualification standard for the item's score point range.

Appendix E: Mathematics Performance Tasks, Reader Training Model

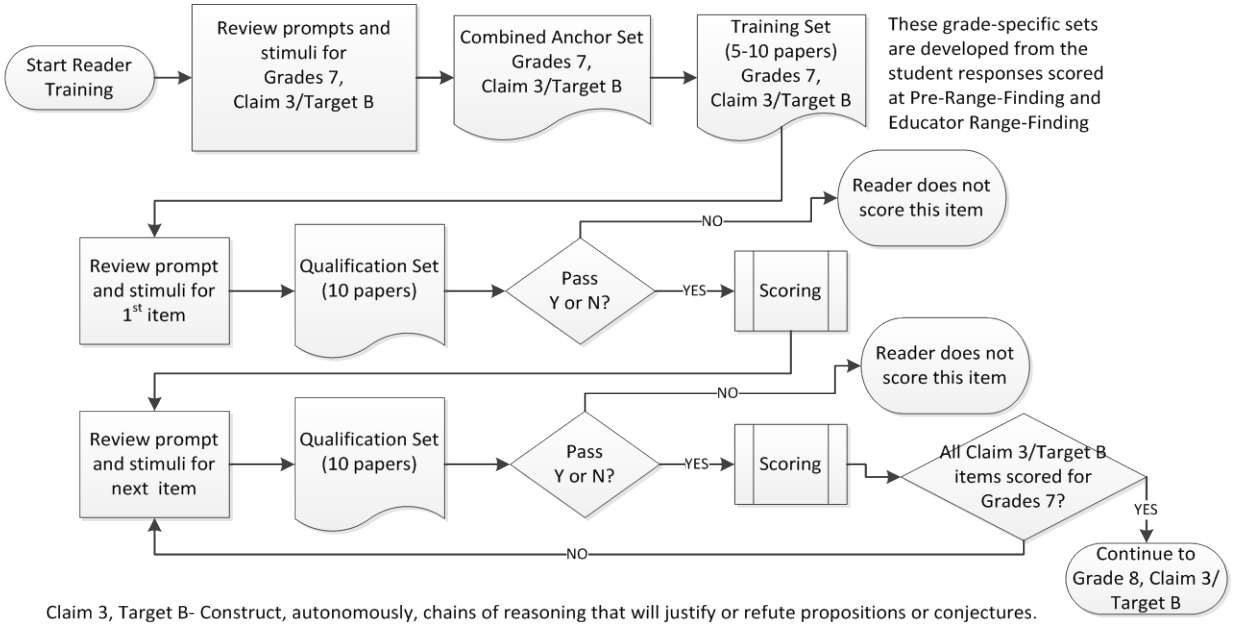


"Pass Y or N?" refers to whether or not the reader meets the qualification standard for the item's score point range.

Appendix F: Mathematics Claim 3 & 4 Short-Text Items, Reader Training Model

Reader Training Model
Example: Claim 3 Short Texts

This model is flexible for Reasoning Items. It is expected that many of these CAT items will require their own anchor/training set.



Claim 3, Target B- Construct, autonomously, chains of reasoning that will justify or refute propositions or conjectures.
Claim 3, Target G- Determine conditions under which an argument does and does not apply.
Claim 4, Target B- Construct, autonomously, chains of reasoning to justify mathematical models used, interpretations made, and solutions proposed for a complex problem.
Claim 4, Target D- Interpret results in the context of a situation.

Process repeats for
Grade 8, Claim 3/Target B
High School, Claim 3/Target B
High School, Claim 3, Target G
Grades 4-6, Claim 3 Target B
Claim 4

"Pass Y or N?" refers to whether or not the reader meets the qualification standard for the item's score point range.

Appendix G: Field Test Condition Codes

Condition Code	Condition Code Category
B	**Blank**
I	**Insufficient** (a) Student has not provided a meaningful response Some examples: • Random keystrokes • Undecipherable text • *I hate this test* • *I don't know, IDK* • *I don't care,* • *I like pizza!* (in response to a reading passage about helicopters) • Response consists entirely of profanity (b) For ELA Full Writes, use "I" code (Insufficient) for responses described above and also if: Student's original work is insufficient to make a determination whether he or she is able to organize, cite evidence/elaborate, and use conventions as defined in the rubrics, or Response is too brief to make a determination regarding whether it is on purpose or on topic
L	**Non-Scorable Language** ELA/literacy: Language other than English Mathematics: Language other than English or Spanish
T	**Off Topic** *for ELA Full Writes Only* Definition: A writing sample will be judged "off topic" when the response is unrelated to the task or the sources or shows no evidence that the student has read the task or the sources (especially for informational/explanatory and opinion/argumentative) • "Off topic" responses are generally substantial responses
M	**Off Purpose** *for ELA Full Writes Only* Definition: A writing sample will be judged off purpose when the student has clearly not written to the purpose designated in the task. • An off-purpose response addresses the topic of the task but not the purpose of the task • Note that students may use narrative techniques in an explanatory essay or use argumentative/persuasive techniques to explain, for example, and still be on purpose • Off purpose responses are generally developed responses (essays, poems, etc.) clearly not written to the designated purpose

For ELA/literacy Full Writes:
If a response is too brief to make a determination regarding whether it is on purpose or on topic, then it should be coded as "Insufficient".

Appendix H: State Letter for Student Alerts (approved for 2013 Pilot)

<DATE>

<CONTACT NAME>

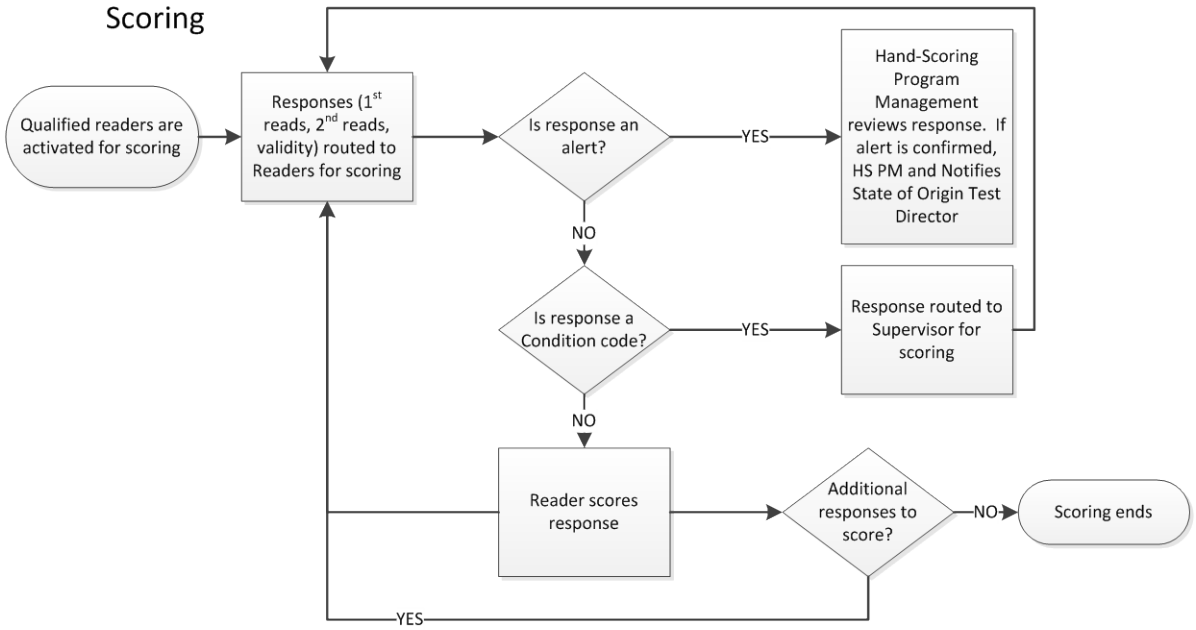
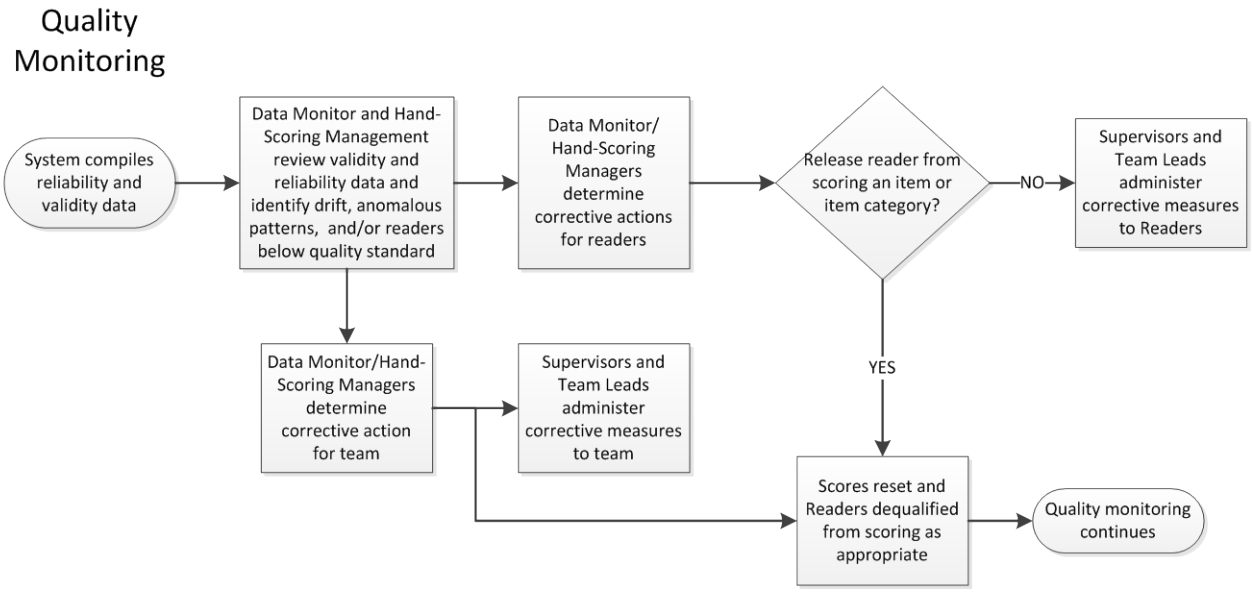
<CONTACT ADDRESS>

Dear <CONTACT>,

As the Scoring vendor for the Smarter Balanced Field Test, we are contacting you on behalf of our professional scoring evaluation staff. The enclosed student response has been identified as one which may require special attention. The paper has been processed according to our established procedures. Please review the content of this document immediately and determine what further steps will be necessary at the school/district level. The information below identifies the source of the document. Thank you for your prompt response and attention to this concern.

Sincerely,

<HANDSCORING MANAGER>
<TITLE>
<PHONE NUMBER>

STUDENT IDENTIFIER: DISTRICT ID: SCHOOL ID: GRADE:

Appendix I: Field Test Scoring and Quality Monitoring Process