

APPENDIX E

MCAS SCORING SPECIFICATIONS

**2018-
2019**



**Massachusetts Comprehensive Assessment System (MCAS)
Scoring Specifications for the 2018-19
MCAS**

Scoring specifications, requirements, and decision rules

Table of Contents

MCAS SCOPE OF SCORING.....	4
SCORING SERVICES STAFFING FOR MCAS	5
SCORER QUALIFICATIONS AND REQUIREMENTS.....	7
SCORER TRAINING.....	8
QUALIFICATION SETS.....	9
THE SCORING PROCESS.....	9
OVERVIEW	9
CONDITION CODES	10
QUALITY CONTROL.....	11
READ-BEHINDS	11
DOUBLE-BLIND SCORING	11
VALIDITY PAPERS.....	12
<i>Validity Paper Insertion.....</i>	<i>12</i>
<i>Validity Paper Selection.....</i>	<i>12</i>
VALIDITY INTERVENTION STANDARDS.....	13
VOIDING SCORER WORK.....	13
SEEDED PAPERS AND EQUATING ITEMS.....	14
SCORER MONITORING REPORTS	14
ITEM BENCHMARKING.....	15
OPERATIONAL BENCHMARKING	15
FIELD TEST BENCHMARKING	15
CRISIS PAPERS.....	16
SCORING FACILITIES	16
SCORING CENTERS	16
SECURITY OF SCORING FACILITIES AND SYSTEMS.....	17
Appendix A.....	18
Appendix B.....	20
Appendix C.....	21
Appendix D.....	22
Appendix E.....	23
Appendix F.....	24
Appendix G	25
Appendix H.....	26

**MCAS
SCORING SPECIFICATIONS, REQUIREMENTS, AND DECISION RULES**

The activities described in this document may be modified based on the mutual decision of the DESE and Cognia.

MCAS Scope of Scoring

The 2018-19 MCAS consists of constructed response and Essay (previously referred to as TBE items) in grades 3-8 and high school

Scope of Work						
Content	Grade					
	3	4	5	6	7	8
ELA	2 OP ES 1 OP CR 2 EQ ES 1 EQ CR 7 FT ES 6 FT ER	2 OP ES 1 OP CR 2 EQ ES 1 EQ CR 7 FT ES 6 FT ER	3 OP ES 2 EQ ES 10 FT ES	3 OP ES 2 EQ ES 10 FT ES	3 OP ES 2 EQ ES 10 FT ES	3 OP ES 2 EQ ES 10 FT ES
Mathematics	4 OP CR 2 EQ CR 10 FT CR	4 OP CR 2 EQ CR 7 FT CR	4 OP CR 2 EQ CR 7 FT CR	4 OP CR 2 EQ CR 7 FT CR	4 OP CR 2 EQ CR 7 FT CR	4 OP CR 2 EQ CR 7 FT CR
STE			4 OP CR 2 OP SA 2 EQ CR 1 EQ SA 18 FT CR 8 FT SA			4 OP CR 2 OP SA 2 EQ CR 1 EQ SA 16 FT CR 8 FT SA
High School						
	Nov. Retest	Feb. Retest	March Retest	Spring		
ELA	4 CR (Reading) 1 WP (Composition)		4 CR (Reading) 1 WP (Composition)	3 OP ES 2 EQ ES 20 FT ES		
Mathematics	6 OP CR 4 OP SA		6 OP CR 4 OP SA	4 OP CR 2 EQ CR 14 FT CR		
Biology		5 OP CR		5 OP CR 10 FT CR (4pt) 16 FT CR (3pt)		
Physics				5 OP CR 10 FT CR (4pt) 16 FT CR (3pt)		

OP = Operational; EQ = Equating; FT = Field Test;
ES = Essay; CR = Constructed Response; SA = Short Answer; WP = Writing Prompt

Scoring Services Staffing for MCAS

Multiple members of the scoring team will be involved in the MCAS project, both from Cognia and Pearson. The table below summarizes the general responsibilities of each staff member.

Cognia	
Position	Description
Senior Vice President of Operations for Assessment Services	Oversees all aspects of operational and scoring-related activities within the division of Assessment Services
Scoring Project Manager	Manages scoring-related activities and deliverables, scheduling of tasks, and monitoring of quality and production.
Assistant Director, Scoring Content	Oversees the work flow and product of Scoring Content Managers and Scoring Content Specialists within the department of Scoring Services.
Assistant Director, Scoring Operations	Oversees Scoring Center Managers and coordinates the operations and logistics for all scoring centers.
iScore Operations Manager	Maintains iScore system and coordinates data deliverables between Scoring Services and Reporting.
Scoring Center Manager	Oversees scoring center logistics, recruitment of contingent workforce, facility requirements and security.
Scoring Content Group Manager	Manages Scoring Content Specialists within a content group and oversees work flow processes to ensure the quality and production of scoring.
Scoring Content Specialist (Math) Scoring Content Specialist (ELA) Scoring Content Specialist (STE)	Supervises the scoring of content within an assigned contract. Responsibilities include facilitating benchmarking [range-finding] meetings and finalizing the selection of all training materials. The Scoring Content Specialist also trains and supervises scoring leadership and monitors the training and scoring of items within a project. The SCS has the overall responsibility of ensuring accurate and consistent scoring according to the approved client guidelines for their content area.
Scoring Supervisor	Scoring Supervisors work under the guidance of a Scoring Content Specialist. They are responsible for training items and ensuring consistency across assigned grades and content. They also respond to any questions that arise during training

	and throughout scoring. They monitor scoring production and quality in iScore.
Scoring Team Leader (STL)	Scoring Team Leaders work under the supervision of Scoring Supervisors and lead a small group of scorers (typically no more than 11 scorers to each STL). STLs are responsible for quality control by maintaining accurate and consistent scoring. They also provide coaching as needed.
Scorers	Scorers review, evaluate, and assign scores to student work based on training and item criteria.
Pearson	
Position	Description
Portfolio Manager	Manages Pearson’s responsibilities across multiple departments
Project Manager (Scoring)	Manages scoring-related activities and deliverables, scheduling of tasks, and monitoring of quality and production.
Scoring Content Specialist (Math) Scoring Content Specialist (ELA)	Supervises the scoring of content within an assigned contract. The Scoring Content Specialist trains and supervises scoring leadership and monitors the training and scoring of items within a project. The SCS has the overall responsibility of ensuring accurate and consistent scoring according to the approved client guidelines for their content area.
Scoring Directors	Scoring Directors work under the guidance of a Scoring Content Specialist. They are responsible for training items and ensuring consistency across assigned grades and content. They also respond to any questions that arise during training and throughout scoring. They monitor scoring production and quality in ePen.
Scoring Supervisors	Scoring Supervisors work under the supervision of Scoring Directors and lead a small group of scorers. Responsible for quality control by maintaining accurate and consistent scoring. They also provide coaching as needed.
Scorers	Scorers review, evaluate, and assign scores to student work based on training and item criteria.

Scorer Qualifications and Requirements

Cognia and Pearson, through its staffing partners, seeks to employ scorers with a wide range of educational and professional experience. Information on demographics and educational background will be collected.

Cognia and Pearson will recruit individuals who meet or exceed the contract-specific requirements to fill scorer and scoring leadership positions. The scoring criteria for MCAS stipulate the following:

Position	Minimum Education Requirement	Other Requirements
Scorer	Grades 3-8: <ul style="list-style-type: none"> • 48 college credits AND • Passed at least 2 college classes related to the content area being scored High School: <ul style="list-style-type: none"> • 4-year college degree AND <ul style="list-style-type: none"> ○ A degree related to the content area being scored OR ○ 2 classes related to the content area being scored and demonstrated scoring experience in the content area 	<ul style="list-style-type: none"> • Must be at least 18 years of age • Cannot be under contract to Massachusetts schools, including as teachers, administrations, and para-professionals
Scoring Team Leader	Grades 3-8: <ul style="list-style-type: none"> • 4-year college degree AND • Passed at least 2 college classes related to the content area being scored High School: <ul style="list-style-type: none"> • 4-year college degree AND <ul style="list-style-type: none"> ○ At least 4 classes related to the content area being scored OR ○ 2 classes related to the content area being scored and demonstrated scoring experience in the content area 	<ul style="list-style-type: none"> • Must be at least 18 years of age • Cannot be under contract to Massachusetts schools, including as teachers, administrations, and para-professionals
Scoring Supervisor	Grades 3-8: <ul style="list-style-type: none"> • 4-year college degree AND • Passed at least 2 college classes related to the content area being scored High School: <ul style="list-style-type: none"> • 4-year college degree AND <ul style="list-style-type: none"> ○ At least 4 classes related to the content area being scored OR ○ Less than 4 classes in the content area with approval from the DESE 	<ul style="list-style-type: none"> • Must be at least 18 years of age • Cannot be under contract to Massachusetts schools, including as teachers, administrations, and para-professionals

Potential scorers will submit documentation, including transcripts and resumes, to verify eligibility to work on the project. If hired, all scoring associates will be required to sign a non-disclosure / confidentiality agreement.

Scorer Training

Scorer training will begin with an introduction to scoring and an overview of the MCAS program. This could include the purpose and goal of the testing program and any unique features of the test and/or testing population. There will also be a general discussion about the security, confidentiality, and proprietary nature of testing, scoring materials, and procedures.

Training materials (in electronic or paper format) will be available to scorers at all times during scoring and will include:

- Student prompt Scoring rubric
- Item Sample response and training notes
- Anchor Set
 - clear examples of typical student responses at each score point
 - Presented in score point order
- Practice Set
 - May include atypical responses or student work that helps demonstrate the cut-points between adjacent score points

Training materials will include examples of all score points whenever possible.

A Scoring Content Specialist or Scoring Supervisor will lead the training for each item. Training may occur in various formats: face-to-face in a training room, through a pre-recorded training module, or through an online training system with audio capabilities which allows for the possibility of remote training. Regardless of the method of training, the approach will follow this sequence:

1. Review of the student prompt, the scoring rubric, associated sample responses, and training notes
2. Review of the anchor set
3. Analysis and discussion of each anchor response, its assigned score and associated, detailed scoring rationale
4. Scoring of a set of practice responses
5. Discussion of each practice response, revealing the actual score assigned to the student response and explaining the scoring rationale
6. Methodical review of all scoring criteria while paying particular attention to the fine lines that determine the cut-points between adjacent score points
7. Question and answer segment addressing any remaining scorer questions
8. Administration of a qualification set consisting of 10 pre-scored responses, scored independently and deployed randomly to each scorer
9. Review of qualification results
10. Start scoring live student responses for the MCAS project (qualified scorers only)

Qualification Sets

Qualification sets are used to ensure that scorers have successfully internalized the scoring standards before they begin scoring each item. Guidelines for the qualification sets are:

- Each qualification set will contain 10 responses.
- Qualification sets are administered through the iScore system and distributed to the scorers unscored and in random order.
- In order to qualify, scorers are required to meet the passing threshold 70% exact agreement and 90% exact plus adjacent agreement. For multi-trait ELA items, the 70/90 passing threshold must be met on each individual trait.
- Scorers who do not pass the first qualification set may be retrained. After retraining, they will take a second qualification set, and must achieve the passing threshold.
- Any scorer who does not pass will not be allowed to score the item. They will either be trained on a different item or dismissed from the scoring project.
- Scoring Team Leaders are required to qualify on each item STLs must pass at a higher threshold and must meet a minimum requirement of 80% exact agreement and 90% exact plus adjacent agreement. For multi-trait ELA items, the 80/90 passing threshold must be met on each individual trait.
- Scoring Content Specialists and Scoring Supervisors who prepare and conduct the training do not qualify on the item. Scoring Supervisors who are working in an edit scoring capacity for the project will train and qualify on all items.
- Responses included in the qualification set must be approved for use by the Scoring Content Specialist or Assistant Scoring Content Specialist. In addition, they must either be part of the MCAS approved benchmarking set or have matching scores that are agreed upon by multiple people, including scoring leadership.

Scoring Team Leaders receive the same training and undergo the same qualification process as scorers. However, STLs may be trained on the items in advance during a separate leadership training. This provides an additional opportunity to absorb the training materials and it prepares them to fulfill their role during scorer qualification.

The Scoring Process

Overview

The scoring of student responses will be conducted through the iScore or ePEN system which displays images that are received through a data transfer from the online computer-based testing platform or through scanned images of paper-based tests. If there are any rendering issues with any paper-based test books, they will be scored by referring to the actual test book and the scores will be manually entered into iScore.

The scoring systems do not display any student or school identifiable information. Security is maintained during scoring through a highly secure server-to-server interface. It ensures that images are only accessible to those who will be scoring each item or to Scoring management. All responses are tracked through a unique booklet number that is matched to the student records during data processing.

Each scoring day will begin with a review of the item. Monday morning reviews, or any other resumption of scoring after a delay, will start with a broader group refresher on scoring the item.

During the course of scoring, scorers may encounter student responses that indicate the possibility of cheating or some type of testing irregularity. Scorers will score this type of student response based on its own merits then refer them to the Scoring Content Specialist and Project Manager for further processing and notification of the DESE. Upon DESE decision, any score change requests can be made prior to final reporting.

Condition Codes

Scoring Services makes every attempt to score each student response. However, when a response does not conform to the score point parameters, scorers have the option to apply the following codes:

MCAS Legacy Tests (High School STE and Legacy Retests in Math, Reading, and Composition)

- **Blank:** No deliberate marks in the answer space
- **Unreadable:** Faint handwriting or otherwise obstructed student response (for PBT tests)
- **Wrong Location:** A clearly legitimate response to another item on the test (Primarily a code for paper-based testing only.)
- **No Score:** Any other response that cannot receive a numeric score. May include an insufficient amount of original text, direct copy of the prompt, artwork or drawings, or simple words that cannot be assessed as an answer to the prompt, or responses written in a language other than English (or Spanish for High School Math).

Next-Gen Tests (Grades 3-8 ELA and Math, Grade 5&8 Science, High School Math and ELA)

- **Blank:** No deliberate marks in the answer space
- **Unreadable:** Faint handwriting or otherwise obstructed student response (for PBT tests)
- **Wrong Location:** A clearly legitimate response to another item on the test (Primarily a code for paper-based testing only.)
- **Non-English:** Response written entirely in a language other than English, or without enough English or numbers to provide a score (except for tests that are translated to Spanish, which will be scored)
- **Off Topic:** Response does not address the topic or task for the item. The response may be irrelevant to the prompt or a statement that the student is refusing to answer.
- **Direct Copy:** Direct copy of the test from the passage or question

Responses that are identified as Unreadable or Wrong Location undergo a separate resolution process. They will be routed to the Scoring Content Specialist or Scoring Supervisor. Responses will be reviewed, and the appropriate score assigned. Furthermore:

- **Unreadable paper responses** will be reviewed by consulting the student's original test booklet or by requesting a re-scan of the student work. If the response can be read through either method, an appropriate score will be assigned. Completely unreadable responses will not receive a numeric score.
- **Wrong Location responses** (limited to paper-based tests) will be reviewed by a Scoring Supervisor or Scoring Content Specialist. Their broader access to iScore allows them to review all student work and assign the correct scores for each item. Wrong locations can only be scored when it is clear that the student was attempting to respond to another question.

All other condition codes have no resolution process and will not receive a numeric score. Legacy Composition Re-tests have a separate No-Score resolution process. Based on No-Score meetings held the last time these re-tests prompts were used, Scoring Content staff will apply the no-score resolution to these prompts as well. Any no-score examples that do not conform to decisions from the No-Score meetings will be sent to the DESE for consultation.

Quality Control

While all scorers must first train and qualify to gain access to scoring student work, they must maintain acceptable levels of accuracy to continue scoring. iScore provides the opportunity to employ multiple quality control tools in order to monitor accuracy and consistency throughout scoring.

Read-Behinds

Read-behind (also known as “Back Reading” in Pearson’s scoring system) scoring allows the STLs and Scoring Supervisors to monitor the performance of each scorer. It provides an immediate real-time snapshot of a scorer’s accuracy and the opportunity to provide individualized counseling as needed.

Read-behinds are conducted throughout the course of scoring by Scoring Leadership. Scorers are not aware of which responses are designated for read-behinds.

If there is exact agreement between the scorer and the STL, no action is taken. If there is a difference in scores, either adjacent or discrepant, the STL score becomes the score of record. The STL may counsel the scorer on the reason for the disagreement.

The number of read-behinds conducted per scorer will vary, as STLs will focus their attention on scorers as needed. The general guideline is for STLs to conduct at least 10 read-behinds per scorer per full day of scoring. However, Grades 3-8 ELA and Math may have fewer read-behinds conducted since these grades also use Validity Papers as the primary method of monitoring scorer accuracy. STLs will conduct more read-behinds on scorers who require counseling or who are at the lower threshold of accuracy, and fewer read-behinds on scorers who have repeatedly demonstrated high accuracy in scoring.

To ensure the accuracy of the STL, Scoring Supervisors and scoring management have the ability to conduct a review of their read-behind work. The Scoring Supervisor has access to all responses that were reviewed and may compare scores to verify the accuracy and consistency of scoring.

Double-Blind Scoring

While read-behinds measure scorer accuracy in relationship to leadership, double-blind scoring provides statistics on scorer-to-scorer agreement. Double-blind scoring is the practice that refers to a method where the same response is routed to two scorers. The response is independently and anonymously reviewed by each scorer. In double-blind scoring, neither scorer knows which response will be (or already has been) scored by another randomly selected scorer.

For MCAS operational scoring, all responses in Grades 3-8 will be scored at a 10% double-blind rate. In High School, all operational responses will be scored at a 100% double-blind rate. For field

test scoring, all responses will be scored at a 10% double-blind rate for all grades and content areas.

If scorers are adjacent in their scoring of a response, the higher score becomes the score of record. If the scorers are discrepant in their scoring, the response will be sent to an STL for arbitration. The STL will review the response, provide the final score, and counsel scorers as needed. The resolution chart (Appendix XX) summarizes how double-blind scores are processed for the purposes of the student’s score of record.

Validity Papers

Pearson will use Validity Papers for all operational and equating items they score (Grades 3-8 Math and ELA). The chart below shows the quality metrics that must be maintained by scorers. Scorers who fail to maintain these specified quality metrics are released from the project and their responses are reset.

MCAS Quality Metrics				
Subject Area	Earned Score Points	Reliability Minimum Perfect Agreements	Validity	Perfect & Adjacent Agreement
ELA /Math Single Trait	3 pt (0,1,2)	80%	80%	96%
	4 pt (0,1,2,3)	70%	70%	96%
	5 pt (0,1,2,3,4)	65%	65%	95%
ELA /Multi Trait	2- traits	65%	65%	96%

Validity Paper Insertion

During the first 2 days of live scoring, validity papers will be embedded on all MCAS ELA items at a rate of 6% and 3% for MCAS Math items resulting in expected validity counts of between 8 and 25 papers per scorer per day. For example, item, MA301188. If scorer X met the expected scoring rate of 60 papers per hour, at an embedded rate of 3% for math items, scorer X would expect to score 12 validity papers within that 6.5 productive hour day.

On the 3rd day of live scoring validity papers will be embedded for MCAS items at a rate of 4% for ELA and 2% for Math items. At these rates, based on the average expected scoring rates and a 6.5-hour productive day, scorers should expect to score a minimum of 6 and upwards of 19 validity papers per day. Using the same item example above, scorer x scoring MA301188 would expect to score 8 validity papers during day 3.

Validity Paper Selection

Validity papers are first identified by Supervisors from live responses during front reading, regular scoring and read-behinds. When a solid validity paper is identified the Supervisor escalates it to a “Proposed” status in the ePEN validity queue. Once in the validity queue, Scoring Directors review the papers. Those papers deemed suitable are then moved from “Proposed” to “Pending” and “Approved”. If the Scoring Director has any questions regarding the caliber of the validity paper they send it to “Pending” status and the Content Specialist “Approves” or declines. At any time, the Client always has ePEN access to review approved validity papers. Any paper the Customer deems inconsistent with their scoring design can be discussed and immediately retired.

Validity Intervention Standards

Warnings are issued to scorers that do not meet minimum validity metrics after a minimum of 10 validity papers. If after an additional 5 validity papers the scorer has not improved, ePEN automatically locks the scorer out and a 10-paper targeted calibration set is administered. The scorer must attain at least 70% perfect agreement and/or 90% perfect plus adjacent agreement on this calibration set to continue scoring the project. If the scorer passes the targeted calibration ePEN is unlocked and the scorer gains admission to continue scoring. The scorer must continue to maintain project standards for validity as validity statistics will continue to be checked every 10 validity papers. If validity falls below standards at any of these subsequent intervals, the Scorer will be released from the project and all scores assigned immediately reset.

The ePEN2 system checks scorer agreement with validity at the predetermined checkpoints based on item type and score point range. The ePEN2 system will prevent a scorer from scoring who falls below the validity expectation at two checkpoints for an item (the first checkpoint issuing the scorer a Warning) until the scorer passes a targeted calibration set for that item. A scorer who does not pass the targeted calibration set will not be permitted to continue scoring that item. A scorer who passes the targeted calibration set will continue scoring. A scorer who passes a targeted calibration set and falls below the validity agreement requirement at a future checkpoint will be automatically locked by ePEN and prevented from scoring on that item.

Quality Control Summary Table

Quality Control Tools	Operational Scoring	Field Test Scoring
Read-Behind Requirements	10 per reader per full day (may be less with validity paper usage)	2 per scorer, per hour
Double-Blind Percentage	10% Grades 3-8; 100% High School	10%
Recalibration Requirements	5 Responses per item per day, starting on the 2 nd day of scoring	n/a
Validity Paper Requirements	ELA: 6% on first 2 days, then 4% starting on day 3 of the item Math: 3% on first 2 days, then 2% starting on day 3 of the item (Gr 3-8 Math and ELA)	n/a
Embedded CRR Requirements	10 per item on day 1 of Scoring (when validity papers are not used)	n/a
Qualification Requirements	Leadership: 80% exact/90% exact plus adjacent agreement Scorers: 70% exact/90% exact plus adjacent agreement	Leadership: 80% exact/90% exact plus adjacent agreement Scorers: 70% exact/90% exact plus adjacent agreement

Voiding Scorer Work

When scorers meet or exceed accuracy standards, they will continue to receive student responses and may continue to score. If a scorer falls below the established accuracy threshold, they will be retrained. Scoring leadership will determine whether a scorer is allowed to resume scoring.

The iScore and ePEN systems allows us to void a scorer's work. If a scorer fails to maintain accuracy standards, their work for the impacted time frame will be nullified, and the student responses will be routed to other qualified scorers for re-scoring.

When Validity papers are used, the voiding decisions will be made consistent with the Validity Intervention Standards and the Validity Quality Metrics table. When Validity papers are not used, the compilation report, which combines the 5 daily recalibration papers with the read behinds conducted during each day, will be the primary tool to determine voiding. Scorers that do not meet a 70% exact / 90% exact plus adjacent agreement on the compilation report will be voided.

Seeded Papers and Equating Items

For equating constructed response and essay items, 200 responses that were scored in a previous test administration will be included during the scoring of the equating items. These responses will be used as part of psychometric analysis to monitor Scorer drift from year to year. For each equating item, 2500 student responses will be scored from the current administration.

Scorer Monitoring Reports

In order to better monitor the accuracy, consistency, and pace of scoring, iScore is designed to generate a variety of reports to allow scoring leadership to monitor all aspects of a complex testing program. These reports show both the overall performance of the scoring project as well as immediate and real time scorer level data. This provides the opportunity to monitor an individual, the group, and the overall project.

STLs and Scoring Supervisors have access to certain reports to ensure quality scoring. Scoring Content Specialists and scoring management have access to all reports in iScore. The following is a summary of the most commonly used reports:

- The **Read-Behind Summary Report** (Appendix XX) shows the total number of read-behind responses per scorer and shows the number and percentage of responses that were exact, adjacent, and discrepant agreement between the scorer and the STL. The report also provides an overall statistical summary of all scorers working on the item. The report has both a daily and a cumulative option.
- The **Double-Blind Summary Report** (Appendix XX) shows the total number of double-blind responses read by a scorer and will note the number and percentages of exact, adjacent, and discrepant scores. The report also provides an overall statistical summary of all scorers working on the item. The report has both a daily and cumulative option.
- The **Embedded CRR Summary Report** (Appendix XX) will show, for each scorer and for either a particular item or across all items, the total number of responses scored, the number of embedded CRR responses scored, and the number and percentage of exact, adjacent, and discrepant scores.
- The **Qualification Statistics Report** (Appendix XX) will list each scorer by name and ID#, identify which qualification sets they took and their pass or fail status for each one.
- The **Summary Report** (Appendix XX) shows each item and the number of student responses to be scored, the number of responses scored, and the number of double-blind scores provided. This data is provided at the item level.
- The **Compilation Report** (Appendix XX) will show, for each scorer, the total number of responses scored, the number of read-behind responses and the number of scored

Recalibration responses (both individually and combined), and the percentage exact, adjacent, and discrepant.

- The **Pearson Scorer Compilation Report** (Appendix XX) will list the combined scorer performance on Validity Papers, Read behinds, and Double-Blind scoring.

Item Benchmarking

Operational Benchmarking

Grade 10 ELA items will require operational benchmarking activities to fill score-point gaps as a result of the field test score distribution. As part of this process, Cognia scoring staff will review student responses from the operational test and select examples of score points that complete the Anchor, Practice, and Qualification set material to include examples of all available score points. The training materials will then be provided to the DESE content staff for review in advance of a tele-conference phone call to provide final approval of the training materials.

Field Test Benchmarking

In advance of benchmarking meetings, Cognia will make the following preparations:

- The Scoring team will review student responses to field test items and assemble a pack for review at benchmarking meetings. Consistent with DESE expectations, the size and composition of the benchmarking packs will follow these guidelines:
 - ELA: For Essay items, packs will include 13-16 suggested anchor papers (depending on grade level) comprised of each available score point and 5 suggested practice papers. Extra packs will include 50-70 papers. These extra responses may or may not be reviewed depending on decisions made at each meeting. Additional practice and qualification papers will be chosen from the extra pack when possible or picked from other submitted responses when needed. For constructed response items in grades 3 and 4, packs will consist of 7-8 suggested Anchor papers and 5 suggested practice papers. An extra set of 25-35 responses will also be included as potential replacement papers. These extra papers may or may not be reviewed depending on decisions made on the item during the meetings.
 - Math: Math packs will contain 40-50 responses in general descending order. All available score points will be represented. There will be no suggested Anchor papers; Anchor papers will be selected from the approved pack after the benchmarking meeting
 - STE: STE packs will include 30-40 papers of all available score points. Anchor papers will be selected from this set during the meeting. Practice papers will also be selected during the meeting from this set. Qualification sets will be selected after the meeting and may or may not be derived from the benchmarking set. If needed, supplemental training papers may be picked after the meeting and submitted for review.
- All benchmarking packs will contain the item prompt, the rubric/scoring guide, and any training notes.

At benchmarking meetings, the Scoring Content Specialist and CD-State Content Specialists will meet with DESE content staff to review the items and the sample student work.

Following the meetings, Scoring staff will create final training materials to include anchor and practice sets, along with qualification sets. These materials will be assembled based on decisions made in the benchmarking meetings to best convey the scoring decisions during training.

For the majority of field test items, the benchmarking packs and subsequent training material will be selected from students that completed the assessment through the computer-based testing platform, since that is the primary test administration platform in most cases and will be the primary testing method moving forward. Some items in science remain paper-only for test administration in 2018-19, and paper responses will continue to be used for benchmarking purposes.

Crisis Papers

Scorers are trained to identify crisis papers (sometimes referred to as Alert Papers). These include responses which indicate a student may present a danger to themselves or others, the student or another child is in danger, there are indications of sexual or physical abuse, or other specific criteria as specified by the DESE

The Scoring Content Specialist is responsible for immediately notifying the Scoring Project Manager when crisis papers are identified. The Project Manager will provide images of the student work along with the student demographic information to the DESE for review and handling.

Scoring Facilities

Cognia and Pearson maintain multiple scoring facilities and may decide to open temporary scoring facilities when needed. When temporary or new scoring facilities are needed, the following criteria are considered:

- Security
- Scorer capacity
- Proximity to a qualified, diverse scoring pool
- Capacity for electronic equipment
- Access to major highways, public transportation, and parking
- Accessibility for individuals with disabilities
- Adequate number of restroom facilities
- Adequate lighting, heating, and air conditioning

While the iScore system is accessible at any of our facilities behind our firewall, iScore management and operations are located at our Dover, NH, facility in support of all scoring centers.

All of our scoring facilities have the technological capacity to allow for training through WebEx, pre-recorded training modules, or other online training methods. Each facility also has training rooms for face-to-face training or re-training of groups of scorers.

Scoring Centers

The chart below shows the currently planned scoring location for each grade and content area. There may be some changes to the locations and shifts as needed for scheduling, capacity, and staffing purposes.

Grade	Operational Scoring			Field Test Scoring		
	Math	ELA	STE	Math	ELA	STE
3	Mesa, AZ	Charlotte, NC		Longmont, CO	Alpharetta, GA; Longmont, CO	
4	San Antonio, TX	Columbus, OH				
5	Virginia Beach, VA		Menands, NY		Longmont, CO	Menands, NY
6	San Antonio, TX	Iowa City, IA				
7		Mesa, AZ				
8		Charlotte, NC	Menands, NY			Menands, NY
High School	Longmont, CO	Longmont, CO	Menands, NY			Menands, NY

Security of Scoring Facilities and Systems

All scoring centers are secured facilities. They are locked and admission is limited to authorized staff. Access to scoring centers is monitored by a security system that only admits staff who were issued an electronic access card. When arriving for their first assignment, all temporary staff must provide identification prior to being issued an electronic access card. This card also serves as Cognia identification card which must be worn at all times while in the building. In order to access the iScore system, all staff must log-on using their individually assigned username and password which are issued and carefully monitored by Cognia. The iScore system can only be accessed at our scoring centers or through restricted VPN access.

Scorers are not permitted to remove any printed, client-confidential materials such as test items or student responses. All paper training material is collected at the end of each day and securely stored on site. Secure paper shred boxes are placed at various locations throughout the facility to ensure the destruction of any client-confidential paper material after the conclusion of a scoring project. For security and environmental reasons, Cognia will use electronic training materials whenever possible. Access to electronic training materials is used in place of paper whenever possible, accessible only through a secure computer system.

Appendix A

CONFIDENTIALITY AGREEMENT – STAFFING AGENCY

Agreement made this ___ day of _____, 20___, by and between Kelly Services, Inc. (“Kelly”) and _____ (“Temporarily Placed Employee”)

Whereas Kelly, pursuant to an agreement with Cognia is placing the Temporarily Placed Employee Cognia to provide services on a temporary basis; and

Whereas the Temporarily Placed Employee may, in the course of performing services for Cognia, have access to Cognia’s confidential, proprietary information; and

Whereas Kelly has agreed to assure that the Temporarily Placed Employee Preserves Cognia’s confidential proprietary information; and

Whereas Kelly wishes to provide Cognia with assurance that Kelly is instructing its employees of the importance of preserving Cognia’s confidential and proprietary information; and

Whereas the Temporarily Placed Employee, as an express condition of being placed to provide temporary services for Cognia, whether remotely or from the business premises of Cognia, agrees to preserve Cognia’s confidential and proprietary information

Now, therefore, in consideration of the mutual promises contained herein, Kelly and the Temporarily Placed Employee agree and stipulate as follows:

1. The Temporarily Placed Employee, whether working remotely or from Cognia’s premises, in the course of performing assigned services, may have access to Cognia’s valuable, proprietary, and confidential business information and trade secrets as well as confidential test materials and results used by Cognia in servicing its clients (including information prepared by Cognia or supplied to Cognia by its clients.) (Collectively, the “Proprietary Information”). Proprietary Information includes information stored at any location, in any form and accessed by the employee at any time and from any location.
2. The Temporarily Placed Employee agrees to review and observe all of guidelines and procedures regarding access to, and use of, Cognia’s data. The Temporarily Placed Employee and Kelly agree to keep confidential all Proprietary Information and to use such Proprietary Information solely in connection with the performance of services for Cognia. Kelly and the Temporarily Placed Employee understand and agree that any and all Proprietary Information is the property of Cognia. The Temporarily Placed Employee will not use for his or her own benefit, or otherwise disclose, any of the Proprietary Information for any purpose other than the performance of their job duties without the prior written consent of Cognia. The Temporarily Placed Employee shall at all times take whatever steps are necessary to protect the confidentiality of the Proprietary Information.
3. Kelly and the Temporarily Placed Employee agree to maintain and protect the confidentiality of any and all Proprietary Information and not to disclose the Proprietary Information to any third party both during and subsequent to the term of the Temporarily Placed Employee’s placement with Cognia.
4. In the event that the Temporarily Placed Employee or Kelly is required by applicable law, regulation, or legal process to disclose any of the Proprietary Information, he or she will, prior to such disclosure, notify Cognia promptly so that it may seek a protective order or other appropriate remedy or, in Cognia’s sole discretion, waive compliance with the terms of this Agreement. In the event that no such protective order or other remedy is obtained, or that Cognia waives compliance with the terms of this Agreement, the Kelly and the Temporarily Placed Employee will furnish only that portion of the Proprietary Information that, pursuant to advice of counsel, he is legally required. Kelly and The Temporarily Placed Employee will exercise all reasonable efforts to obtain reliable assurance that confidential treatment will be accorded such Proprietary Information.

5. Upon completion of service as for Cognia and prior to his or her departure, the Temporarily Placed Employee agrees to return to Cognia all copies (in whatever form, including electronic media) of any Proprietary Information to which the Temporarily Placed Employee had access while performing services for Cognia.
6. Notwithstanding anything to the contrary in this agreement, the Temporarily Placed Employee's obligations as to the Proprietary Information shall not apply to any portion of the Proprietary Information: (i) that is presently, or will become publicly available or a matter of public knowledge other than by a breach of this Agreement by the Temporarily Placed Employee; (ii) that is lawfully received by the Temporarily Placed Employee from a third party who is not, or was not, bound in any confidential relationship or obligation to Cognia; (iii) that is disclosed with the express written permission of Cognia; or (iv) that is independently conceived by the Temporarily Placed Employee without reference to such Proprietary Information.
7. If any provision of this Agreement shall be determined to be illegal and unenforceable by any court of law, the remaining provisions shall be severable and enforceable in accordance with their terms.
8. The Temporarily Placed Employee and Kelly acknowledge and agree that Cognia may not have an adequate remedy in the event the Temporarily Placed Employee and/or Kelly breaches this Agreement. The Temporarily Placed Employee and Kelly agree that Cognia, in addition to any other available rights and remedies shall be entitled to an injunction restraining the Temporarily Placed Employee and/or Kelly from committing or continuing any violation of this Agreement. The Temporarily Placed Employee and/or Kelly will be obligated to reimburse Cognia its reasonable attorney fees incurred to obtain such injunctive relief.
9. In addition, the parties recognize that Cognia is obligated to protect student information that may have been disclosed in the assessment process. The Temporarily Placed Employee may have access to student related materials that must be held in strictest confidence. ("Student Information"). This Student Information includes, but is not limited to, writing assessment prompts, essays, and the open-ended responses prepared by students of Cognia's clients.
10. The Temporarily Placed Employee and Kelly agree (1) to treat all prompts, essays, and open-ended test items and responses and scores from all test items as confidential; (2) to make no attempt to identify the persons who wrote the responses or the schools that they attended; and (3) to refrain from commenting on these responses except during the reading itself, and then only without referring to individual students. Student Information is not to be reproduced in any manner.
11. It is expressly agreed and understood that Cognia is an intended third party beneficiary of this Agreement and has the right to enforce its provisions.

Kelly Services

By: _____ Title: _____

Temporarily Placed Employee _____

Agreement dated as of _____

Appendix B

MCAS Resolution Charts

Read-Behind Scoring ¹		
Scorer Score	Scoring Supervisor/STL Score	Final
4	4	4
4	3	3
4	2	2

¹In all cases, the leadership score is the final score of record.

Double-Blind Scoring ¹			
Scorer #1	Scorer #2	Scoring Supervisor/STL Resolution	Final
4	4	-	4
4	1	2	2
0	1	-	1
2	4	3	3
1	2	-	2
2	0	2	2

¹ For adjacent scorer scores, the higher score is the final score of record. If a resolution score is needed, as in the case of discrepant scores, the Scoring Supervisor/STL score is the final score of record.

Appendix C

Final Score Determination for MCAS Responses

In cases where multiple scores are provided to the same response, the final score will be determined based on the following hierarchy:

Ranking	Score Source	Provided by:	Notes
1	Edit Score	Scoring Content Specialist, Edit Scoring Supervisor, or entered in iScore/ePen by administrator after scoring paper booklet	
2	Arbitration Score and Read-Behind Score Both Provided	Scoring Team Leader/Scoring Supervisor	If read-behind and arbitration are both provided, the higher score is the score of record
2	Arbitration Score (No Read Behind Perform)	Scoring Team Leader/Scoring Supervisor	Arbitration score is score of record
3	Read-Behind Score (No Arbitration Performed)	Scoring Team Leader/ Scoring Supervisor	If Read-behind score is provided by 2 STLs, the higher read-behind score is the score of record
4	Scorer's Score	Scorer	If a response is scored by 2 scorers, the higher score is the score of record

Appendix D

Read-Behind Summary

Choose Response Code: Choose Time Frame:

Submit

Export To Excel

Contract: -----

Grade: -----

Content: -----

Name	ID#	Scored	Read Behind							
			Total RB	% RB	Exact	% Exact	Adj	% Adj	Disc	% Disc
	17112	227	19	8.4	14	73.7	5	26.3	0	0
	19537	163	16	9.8	14	87.5	2	12.5	0	0
	18034	266	15	5.6	12	80	3	20	0	0
	21212	163	15	9.2	12	80	3	20	0	0
	20855	365	19	5.2	18	94.7	1	5.3	0	0
	21239	443	18	4.1	15	83.3	3	16.7	0	0
	21343	426	18	4.2	15	83.3	3	16.7	0	0
	19556	213	16	7.5	15	93.8	1	6.3	0	0
	19832	341	18	5.3	16	88.9	1	5.6	1	5.6
	18104	305	15	4.9	13	86.7	2	13.3	0	0
	19545	385	20	5.2	16	80	3	15	1	5
	19419	255	17	6.7	16	94.1	1	5.9	0	0
	15836	379	19	5	18	94.7	1	5.3	0	0
	Total	3931	225	5.7	194	86.2	29	12.9	2	0.9

Appendix E

Double-Blind Summary

Reader Double Behind Summary
Content: --

Contract: -----

Grade: -----

Choose Response Code: RDCC025 Choose Time Frame: 08/27/201E

Submit

Export To Excel

Name	ID#	Score d	Double Blind							
			Total DB	% DB	Exact	% Exact	Adj	% Adj	Disc	% Disc
	17112	227	22	9.7	17	77.3	5	22.7	0	0
	19537	163	15	9.2	15	100	0	0	0	0
	18034	266	25	9.4	19	76	6	24	0	0
	21212	163	15	9.2	11	73.3	4	26.7	0	0
	20855	365	29	7.9	23	79.3	6	20.7	0	0
	21239	443	35	7.9	29	82.9	6	17.1	0	0
	21343	426	49	11.5	43	87.8	5	10.2	1	2
	19556	213	24	11.3	24	100	0	0	0	0
	19832	341	31	9.1	23	74.2	8	25.8	0	0
	18104	305	32	10.5	28	87.5	3	9.4	1	3.1
	19545	385	37	9.6	33	89.2	4	10.8	0	0
	19419	255	17	6.7	12	70.6	5	29.4	0	0
	15836	379	30	7.9	26	86.7	4	13.3	0	0
ZZ Total	99999 9	3931	361	5.3	303	83.9	56	15.5	2	0.6

Appendix F

Scorer/Item Qualification Summary

Reader/Item Qualification Summary

Contract:

Grade: -----

Content: -----

Choose Qualification Type

Display Reader Names

Submit

Export To Excel

Name	ID#	R D C C 0 2 5	R D C C 0 2 6	R D C C 0 2 7	R D C C 0 2 8	R D C C 0 2 9	R D C C 0 3 0	R D C C 0 3 1	Total Comple ted CRRs	# Pas s	# Fai l	% Pas s
	Total Passed	15	4	17	18	4	4	12	74			
	Total Failed	2	0	1	0	0	0	1	4			
	Total Number	17	4	18	18	4	4	13	78			
	59806	P/9	P/1 0	P/1 0	P/1 0	P/1 0	P/8	P/8	7	7	0	100
	18498			P/7	P/1 0				2	2	0	100
	21056							P/8	1	1	0	100
	20904			F/3	P/9				2	1	1	50
	17112	P/9							1	1	0	100
	17030							P/8	1	1	0	100
	15567			P/1 0	P/9				2	2	0	100
	21185							P/7	1	1	0	100
	15555			P/9	P/1 0				2	2	0	100
	17411	P/9	P/1 0				P/1 0		3	3	0	100
	19537	P/10							1	1	0	100
	16827							F/6	1	0	1	0
	17130							P/1 0	1	1	0	100
	17099			P/8	P/9	P/1 0			3	3	0	100
	22028							P/8	1	1	0	100
	21401			P/1 0	P/1 0				2	2	0	100
	20031			P/7	P/1 0				2	2	0	100
	18034	P/9							1	1	0	100
	20867			P/9	P/1 0				2	2	0	100

Appendix G

Compilation Report

Compilation Report Contract: 1225 ReadBK2 Grade: 06 Content: Reading

Choose

Response Code: RDCC045 Choose Date: 9/26/2017 Location: All

Message:

Submit

Export To Excel

Void	Reader Name	MPID	Scored	Recal	Exact Recal	Read Behind	Compilation	Exact	%Exact	Adj	%Adj	Disc	%Disc	%Exact + %Adj
Review		67571	0	5	1.0	0	5	1.0	20.0	4.0	80.0	0.0	0.0	100.0
Review		65166	147	5	4.0	10	15	10.0	67.0	5.0	33.0	0.0	0.0	100.0
Retrain		60890	113	5	3.0	12	17	12.0	71.0	5.0	29.0	0.0	0.0	100.0
Retrain		66826	120	5	3.0	12	17	12.0	71.0	5.0	29.0	0.0	0.0	100.0
Retrain		65793	188	5	3.0	11	16	13.0	81.0	3.0	19.0	0.0	0.0	100.0
		65532	111	5	4.0	12	17	14.0	82.0	3.0	18.0	0.0	0.0	100.0
		60751	244	5	5.0	12	17	14.0	82.0	2.0	12.0	1.0	6.0	94.0
		80231	149	5	5.0	12	17	14.0	82.0	2.0	12.0	1.0	6.0	94.0
		80264	145	5	5.0	13	18	15.0	83.0	3.0	17.0	0.0	0.0	100.0
		64851	139	5	5.0	13	18	15.0	83.0	3.0	17.0	0.0	0.0	100.0
		66712	125	5	5.0	13	18	15.0	83.0	3.0	17.0	0.0	0.0	100.0
		66311	297	5	5.0	15	20	17.0	85.0	3.0	15.0	0.0	0.0	100.0
		60737	114	5	4.0	10	15	13.0	87.0	2.0	13.0	0.0	0.0	100.0
		66051	181	5	5.0	12	17	15.0	88.0	2.0	12.0	0.0	0.0	100.0
		80082	151	5	5.0	11	16	14.0	88.0	2.0	13.0	0.0	0.0	101.0

Appendix H

Pearson Compilation Report

Daily Scorer Compilation

09/04/14	ItemName	Employee D	Score Count	Validity Count	Validity Agree %	Validity Adj %	Validity Agree + Adj	VaDisc	Validity Disc %	Double Blind Count	Double Blind Agree %	ReliabilityAgreeAdjacentCount	Double Blind Agree + Adj %	Double Blind Disc %	Read Behind Agree %	Read Behind Agree + Adj %	Read Behind Disc %	
	Item1	1040193	149	14	71.4%	14.3%	85.7%	2	14.3%	50	56.0%		48	96.0%	4.0%			
	Item2	887598	204	16	62.5%	31.3%	93.8%	1	6.3%	84	75.0%		84	100.0%	0.0%	80.6%	100.0%	0.0%
	Item1	877068	306	24	66.7%	29.2%	95.8%	1	4.2%	134	79.1%		134	100.0%	0.0%			
	Item1	1425345	30	4	50.0%	50.0%	100.0%	0	0.0%	0			0					
	Item2	961460	280	24	54.2%	45.8%	100.0%	0	0.0%	152	62.5%		143	94.1%	5.9%			
	Item2	2419150	154	12	75.0%	25.0%	100.0%	0	0.0%	44	56.8%		44	100.0%	0.0%			
	Item1	1425345	281	22	77.3%	22.7%	100.0%	0	0.0%	460	73.3%		457	99.3%	0.7%	87.8%	99.2%	0.8%
	Item2	973935	496	38	81.6%	18.4%	100.0%	0	0.0%	402	68.9%		400	99.5%	0.5%			
	Item2	852890	298	24	83.3%	16.7%	100.0%	0	0.0%	480	80.8%		475	99.0%	0.1%	90.3%	100.0%	0.0%
	Item2	852890	576	44	84.1%	15.9%	100.0%	0	0.0%	468	76.3%		466	99.6%	0.4%	91.8%	95.3%	4.7%
	Item1	2384798	332	26	84.6%	15.4%	100.0%	0	0.0%	108	66.7%		108	100.0%	0.0%	98.5%	100.0%	0.0%
	Item1	843920	266	22	86.4%	13.6%	100.0%	0	0.0%	190	71.1%		190	100.0%	0.0%			