APPENDIX E

Scoring Specifications

for the 2016-17 MCAS

2017

Massachusetts Comprehensive Assessment System (MCAS)

[Scoring Specifications for the 2016-17 MCAS]

SCORING SPECIFICATIONS, REQUIREMENTS, AND DECISION RULES

Table of Contents

MCAS

SCORING SPECIFICATIONS, REQUIREMENTS, AND DECISION RULES

The activities described in this plan may be modified based on the mutual decision of the Massachusetts ESE and Measured Progress.

MCAS Scope of Scoring

The 2016-17 MCAS test consists of Math and ELA items at grades 3-8 and High School, Science items in Grades 5, 8 and High school, and retest opportunities for high school assessments. In grades 3-8 ELA and Math, the test includes a combination of MCAS specific items and PARCC licensed items. Tests may include Common items, Matrix Equating Items, and Matrix field test items. Grade 3-8 ELA and Math tests are sometimes referred to as "Next Gen MCAS" while the science and high school assessments are sometimes referred to as "Legacy Tests" or "Legacy MCAS".

Measured Progress will score all high school and science assessments, as well as conduct all field test scoring. Pearson will score all operational Grade 3-8 ELA and Math assessments. Measured Progress and Pearson will work closely with each other to ensure consistent and seamless scoring between our scoring operations.

CATEGORY	DESCRIPTION
Scorer Hiring and Training	

Scorer Qualifications and Demography	• Measured Progress and Pearson actively seek a diverse scoring pool and typically employs Scorers (also referred to as readers) with a broad range of backgrounds: teachers, business professionals, graduate school students, retired educators, and the like.
	• To ensure this diversity, demographic information such as gender, race, educational background, etc., will be collected.
	• All recruited Scorers will meet, at a minimum, the following requirements:
	  • For Grades 3-8, all Scorers will have at least 48 college credits, with at least two classes related to the content area being scored. Scoring Team Leaders (STLs) and Scoring Supervisors will have a 4-year degree with at least two classes related to the content area being scored.
	  • For High School, the following requirements will be applied:
	  ▪ Scorers must have a 4 year college degree and one of the following
	  • 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
	  ▪ STLs must have a 4 year college degree and one of the

	following: • 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 ▪ Scoring Supervisors must have a 4 year college degree and one of the following: • At least 4 classes related to the content area being scored or • Less than 4 classes related to the content being scored with approval from the ESE • In addition, screened bilingual applicants must be proficient in English and Spanish. An applicant must be able to speak, read, write, and translate to and from English and Spanish. An applicant must be able to carry out the job responsibilities listed above in both English and Spanish. • Potential Scorers will submit documentation (resume and/or transcripts, etc.) as evidence of meeting the education and experience requirements. • At the end of the program year, Measured Progress will provide Massachusetts with a report that summarizes the qualifications and demographics of all MCAS Scorers by scoring site and subject area in the annual technical report. Demographic information captured will include standard categories such as educational background, gender, etc. • All Scorers must be at least 18 years of age, regardless of any advanced educational degrees. No employees under contract to Massachusetts schools, including teachers, administrators, and para-professionals, will be recruited to work on the MCAS contract • All Scorers will sign a non-disclosure/confidentiality agreement *(Appendix A)*.

Scorer Training	• Scorer training will commence with an overview of the MCAS program. This will include the purpose and goal of the testing program and any unique features of the test and/or testing population. • There will be a general discussion about the security, confidentiality, and proprietary nature of testing, scoring materials, and procedures. • Item specific training may occur in a variety of ways: o Live, in-person training in a conference room or other area well suited for discussion o Live training through WebEx or other types of communication system. The trainer may be on site or training remotely. o Pre-recorded training modules • All scoring personnel will undergo training that focuses on the features of the MCAS Scoring Rubrics. Scoring Leadership will review its criteria to help ensure consistency in understanding the content standards. The progression of each criterion's language on the rubric will be discussed. Anchor papers will be used as exemplars for a variety of possible score

	points. Scorers will be trained to evaluate each response as a focused unit. • For each scoring item, Scorers will thoroughly review and discuss the associated MCAS Scoring Rubric. • Scorer training will involve careful reviews of actual student responses, organized into training materials. • Scorers will review actual responses with an item- specific Anchor Set, and they will be instructed to refer back to the Anchor Set frequently during all scoring activities. • Anchor paper responses are all approved by the ESE and cannot be changed without the consent of the ESE. • Anchor responses will be presented in a pre-determined order. • The Trainer (Scoring Content Specialist/Scoring Supervisor) will announce the anchor response score and explain the scoring of the response in relation to the score-point description and Training Notes, allowing Scorers to internalize typical characteristics of each score point. • Practice papers may contain responses that are more unusual and/or less solid (e.g., are shorter than normal, employ atypical approaches, contain either very low and very high attributes, or written in hard-to-decipher handwriting). • During the review of the practice material, the Trainer will often focus review efforts on the lines between two score points or other scoring issues that are traditionally difficult for Scorers to internalize. • Scorers will independently read and score the response; the Trainer will then indicate the actual score, and explain the scoring rationale for the response. • Scorers will have access to all training material, either with paper copies, or access to electronic versions of the material. • The training materials will include examples of all score points when possible. • To ensure scoring leadership, one Scoring Content Specialist will be identified as the primary contact person for scoring per content area. • The Scoring Content Specialists will have the overall responsibility of making sure that Scorers score accurately and consistently, according to the approved scoring guidelines. • Measured Progress's Scoring Supervisors and Scoring Team Leaders (STLs) will play a major role in Scorer training, re-training, and the monitoring of Scorers throughout the training and scoring process. The Scoring Content Specialist will closely supervise scoring leadership personnel and will be ultimately responsible for all training and scoring activities. • For PARCC licensed items that are included in the MCAS test, items will be scored by PARCC Scorers following the same rules and procedures for the PARCC test.
Item Benchmarking	
Operational Benchmarking	• For Legacy Grade 10 Writing items, operational benchmarking will occur to select training materials from the Massachusetts students

	taking the test. In advance of scoring, Scoring staff will select benchmarking packs of student work and send the packs to the ESE content staff for review. In meetings, the responses will be discussed to finalize the scoring rules for the item and approve the training material for the item. • For operational items, supplemental training materials may be created. Additional Practice papers and qualification papers may be selected from scored field test responses to add to the training materials to reflect trends in answers noticed during field test scoring. These expanded training materials will be submitted for ESE review and approval in advance of scoring.
Benchmarking Process	• The 2016-17 MCAS contains field test (FT) items in all content areas in grades 3-8 (5&8 for science), as well as High school Biology, Physics and Technology & Engineering. • For field test scoring, Measured Progress will score 1800 student responses per item with at least 10% of the responses scored by multiple Scorers (double-blind scoring). Unreadable and Wrong Location responses will not be resolved for field test items unless a small population is taking the item, and all student data is needed for analysis. The process of benchmarking the 2016-17 field test responses will involve three distinct processes, as described below: • Measured Progress Scoring Services staff will review unscored responses that have been uploaded into iScore and select student responses that represent the entire score point range, arranged in descending order. This pack will include the item, rubric, a suggested anchor pack, followed by a pack of extra responses. The packs presented in the benchmarking meeting will not have scores or annotations on them (Scores and annotations can be added at ESE request). Final scores are decided in the meeting by the ESE staff. • The benchmarking committee will consist of content staff from Scoring, CDD, and the ESE. • The committee will review all open response items. Final Anchor paper decisions will be made during the meeting. • Following benchmarking meetings, Measured Progress Scoring Services staff will create training material, using the ESE approved scores. All training material can be provided to the ESE. • Consistent with ESE expectations, the size of the benchmarking packs will follow these guidelines: o Text Based Evidence (TBE) item types will include 75-100 responses and an additional 12-13 Anchor Papers. Practice papers will be chosen from the extra set and will include 4-12 papers and contain at least one of each score point (when available). A Qualification set of 10 papers will be chosen from the benchmarking set when possible, but may be supplemented after the meeting if needed. o Open Response (OR) item types will include 30-45 responses

	with an additional 12-13 Anchor papers. Practice papers will be chosen from the extra set and will include 4-12 papers and contain at least one of each score point (when available). Qualification set of 10 papers will be chosen from the benchmarking set when possible, but may be supplemented after the meeting if needed. o TBE items will include discussion and training of the Conventions trait through the same training material used to train the Idea and Development Trait. • In addition to the embedded field test items, there will be a separate "tryout" test for additional items in Grades 5 and 8 Science. These items will follow the same process as field test scoring, but occur in a different testing and scoring widow.
Qualification Sets	• Scoring leadership uses Committee Reviewed Responses (CRRs) as Qualification Sets to determine if Scorers are calibrated to the scoring standard before they are allowed to score each item. o For TBE and OR items, Qualification sets consist of 10 previously scored field test responses that are selected as qualification criteria, or from unscored responses that are agreed upon by multiple members of the Scoring Staff. o Qualification Sets are entirely pre-selected before scoring begins. o Qualification Sets are given as a set. **<u>Selection Criteria</u>** • To qualify as a Qualification Set, each response must meet the following criteria: o the score must contain a mid-range score-point example; o a Scoring Content Specialist or Assistant Scoring Content Specialist has approved using the response in a CRR Set. • If a Scorer fails to meet the passing threshold on the second Qualification Set, Scoring Department leadership will release the Scorer from scoring that item. **<u>Administration Procedure</u>** **Leadership** • Scoring Supervisors and STLs will have their own leadership training session preceding Scorer training. Scoring Supervisors and STLs who enter scores of record will have a higher standard on Qualifying Sets: a minimum accuracy-scoring rate of 80% exact and 90% adjacent agreement. **Scorer** • **Scorers may have <u>two</u> opportunities to qualify for scoring each scoring item.** If Scorers are unable to attain a score match of at least 70% exact/90% adjacent agreement on the first qualifying set, they will be retrained by discussing the Qualification responses in terms of the score-point descriptions and the original Anchor Set. Following this training,

<table>
<tr><td></td><td>Scoring Department leadership will administer a second set Qualification Set. If they achieve a scoring accuracy rate of at least 70% exact/90% adjacent agreement on either of the Qualification Sets, Scoring Department leadership will allow them to score student responses.
• Bilingual Scorers (for high school math) will train and qualify on the English prompt. After qualifying, they will be eligible to score both English and Spanish responses.
• For the high school legacy Writing test, Scorers must meet the qualification standard on each trait in order to be qualified to score.
• For PARCC licensed items, scorers will qualify under PARCC guidelines and score the items after becoming PARCC qualified.</td></tr>
<tr><td>Scoring</td><td></td></tr>
<tr><td>Scoring Parameters</td><td>• The MCAS Scoring scale options include, depending on the content and the item type—0, 1, 2, 3, 4, 5, or 6. For the Legacy Grade 10 Composition test, the essays are scored on two traits; a Standard English Conventions trait scored within a 1-4 point range, and a Topic Development trait scored within a 1-6 point range. For TBE items in Grades 3-8, responses are scored on two traits; an Idea Development trait scored within a 1-4 point range (Grades 3-5) or 1-5 point range (Grades 6-8) and a Conventions trait scored on a 1-3 point range.
• Although the Scoring Department will make every attempt to score every page of each response, when a response does not conform to score point parameters (including 0), Scorers have the choice of designating the following options:

Legacy MCAS Test (Grades 5&8 Science; All High School Tests)
o Blank: No attempt to respond to the item and may include stray marks.
o Unreadable: Faint monitor image, tiny/light handwriting, poor penmanship, etc.
o Wrong Location: A legitimate response to another item
o No Score: May include a variety of situations, including
o There is an insufficient amount of original text from the student to accurately apply the scoring guide. This would include all or almost all of the response being a direct copy of the prompt, artwork or drawings, or a simple work or words that cannot be assessed.
o The response is completely unrelated to the given prompt.
o Any response in a language other than English (Except Spanish in high school math).
o Any other non-scorable case as defined by the ESE.

Resolution Process
<table><tr><td>• Blank</td><td>• Responses scored Blank will be sent to another Scorer for a second read. Responses scored Blank twice are converted to zeros (‘0’s) for reporting purposes. Should a discrepancy arise,</td></tr></table></td></tr>
</table>

	this will be resolved by the STL/Scoring Supervisors as shown in the Resolution Chart (*Appendix B*).
• **Unreadable**	• Those responses judged unreadable will be forwarded to an edit Scoring Supervisor who is specially trained to ensure that students receive all points they deserve. If needed, they will be resolved by consulting the original test booklet.
• **Wrong Location**	• A legitimate response to another item. Those responses judged Wrong Location will be forwarded to a specially trained edit Scoring Supervisor for response appraisal and scoring and assigning the score to the appropriate item.
• **No Score**	• No Score responses do not receive a numeric score. No score responses in High School Legacy Writing are reviewed at No-Score meetings after scoring for final resolution and classification.

- No-Score resolution for the Writing (Composition) Legacy test will follow the same procedures as past administrations
    - A representative sample of No-Score responses will be assembled and reviewed at a meeting between the ESE and Scoring Content staff.
    - During the No-Score meeting, the responses will be classified as Insufficient Evidence, Non-English, Off Topic, Plot Summary
    - After the meeting, the No-Score decision rules will be applied to all responses initially designated as “No score” and will be exported for reporting.

Next Generation MCAS (Grades 3-8 ELA and Math

MCAS Condition Codes	PARCC Condition Codes	Next Gen MCAS Description
BL-Blank	A-No Response	No attempt to respond to item Either no or stray marks on paper
UN-Unreadable	B- Response is Unintelligible or Undecipherable	Response cannot be read because of poor penmanship, spelling that cannot be deciphered, writing is too small, too faint to see, or only partially visible
NE-Non-English	C – Response is Not Written in English	Response written entirely in a language other than English, or without enough English/numbers to provide a score
OT-Off Topic	D – Off Topic E – Refusal to Respond	Response does not address the topic or task for item Response is irrelevant to the prompt, but could be a reasonable answer to a different prompt Response states that student is refusing to participate in testing
Score 0 (math)	F – Don't Understand/Know	PARCC items scored under this condition code will be scored as a zero and not a condition code
DC-Direct Copy (P-Repeats Prompts)	No PARCC code	Direct copy of test from the passage or question

Resolution Process

• **Blank**	• Responses scored Blank will be sent to another Scorer for a second read.
• **Unreadable**	• For paper tests, those responses judged unreadable will be forwarded to an edit Scoring Supervisor who is specially trained to ensure that students receive all points they deserve. If needed, they will be resolved by consulting the original test booklet.
• **Wrong Location**	• A legitimate response to another

<table>
<tr><td></td><td><table>
<tr><td></td><td>item. Those responses judged Wrong Location will be forwarded to a specially trained edit Scoring Supervisor for response appraisal and scoring and assigning the score to the appropriate item. No response will receive a wrong location as a final score – this is an intermediate code only to allow scoring leadership to properly score.</td></tr>
<tr><td>• All other condition codes</td><td>• All other condition codes will not receive a numeric score.</td></tr>
</table></td></tr>
<tr><td>Alert Papers</td><td>• Scorers are trained to identify “alert” papers (i.e., student responses indicating possible danger to self or others and any other objectionable responses specified by the ESE and immediately refer them to the Scoring Content Specialist.
• The Scoring Content Specialist will be responsible for notifying the Scoring Project Manager, who will coordinate with Client Services.
• MCAS program management will notify the ESE and provide the student and school information.</td></tr>
<tr><td>Number of Readings</td><td>• In grades 3-8, all operational responses will be ten percent (10%) double blind scored (see Double Blind Scoring). For High School, all operational responses will be one-hundred percent (100%) double blind scored.
• A chart outlining the rules for handling adjacent and discrepant scores is attached as Appendix B.</td></tr>
<tr><td>Scoring Methodology</td><td>• Scorers will score almost all student responses from computer images (the only exceptions are for responses that we are unable to scan into our scoring engines). All scoring is blind - no student names are visible to Scorers. Booklet numbers within the scoring system link student responses.
• Each scoring day will begin with an individual or group review of the item/assignment, the scoring rubric and training notes, Anchor Set, and other key reference responses. Monday morning reviews (or after any break of two or more days) will begin with a group review and discussion of the same.
• Measured Progress and Pearson will maintain security during scoring by using a highly secure, server-to-server interface to ensure that access to all student response images is limited to only those who will be scoring or to those who work for Measured Progress or Pearson in a scoring management capacity.
• Any damaged answer/test booklet will be hand-scored. This data will be entered into the scoring system. Score files will be merged with student demographic information by Data and Reporting Services.
• Any student response indicating potential cheating and/or security lapses before, during, or after the test administration will be scored based on the merits of the response and then forwarded to the Scoring Content Specialist for review. If further attention is warranted, the response will</td></tr>
</table>

	be forwarded to the Scoring Project Manager and MCAS Program Management.
Scorer Monitoring Overview	• During scoring, the scoring system will enable a constant measuring and monitoring of Scorers for scoring accuracy and consistency. Each Scorer's reading rate and total number of scored responses will also be monitored. • Following qualification, Scorers will be required to demonstrate their ability to score student responses accurately and consistently throughout the training, qualification, and scoring processes. Scorers will be required to maintain an acceptable scoring accuracy rate (70% exact/90% adjacent agreement) on a daily basis as measured by: o Read-behinds o Double-blinds o Embedded Responses (when Validity sets are not used) o Recalibration Sets (when Validity sets are not used) o Compilation reports (a combination of Recalibration assessments and read-behinds) o Validity Papers ➢ When Validity Papers are used (All PARCC items, and NextGen MCAS if ESE requests), the performance on daily Validity Papers will be the primary measure to determine scorer performance for the day. When Validity Papers are not used, the combination of Recalibration sets and Read Behinds, as measured on the combined Compilation Report, will be the primary measure to determine scorer performance for the day. Read Behinds and Double Blind scoring is conducted in all cases. • If a Scorer falls below standard on a particular item, the Scorer may have an opportunity to be retrained on that particular item and, pending approval by scoring leadership, be allowed to resume scoring. • If scoring is resumed, the Scoring Supervisor/STL will increase the number of read-behind scorings for the individual and will continue to do so until the Scorer is consistently above the accuracy standards. • Scorers will have only two opportunities to be retrained on a particular item. If they fall below standard for the third time based on data collected on a compilation report, they will be dismissed from scoring that item. • At different times during each day of scoring, either individually or as a group, the Scoring Content Specialist and Scoring Supervisor/STL will provide points of clarification about the scoring of specific item and key elements of the rubric. • Scoring rules are in place to determine the score of record when a student is scored by two Scorers, or when a score is provided by Scoring Leadership. (see *Appendix C*)
Recalibration Sets	• Commencing the second day of scoring for each item, Scorers will take one (1) Recalibration Set each day to determine that they remain calibrated to the scoring standard. • Recalibration Sets will consist of five (5) pre-scored responses representing a variety of possible scores.

	• Recalibration Sets will be used to measure whether a Scorer is meeting the standard for scoring MCAS items – specifically, the standard calls for at least 70% exact/90% adjacent agreement on that Recalibration Set. • The results of the Recalibration set will be combined with Read Behind data for the day in a Compilation Report to judge the Scorer's performance on an item each day. **Selection Criteria** • Scoring Leadership selects Recalibration Sets from recently scored responses. • To qualify as a Recalibration Set response, each response must meet the following criteria: o the response must have been given the same score by at least two different scoring bodies (i.e., Scorers, STLs, Scoring Supervisors, and the Scoring Content Specialist) or it must have been derived from the item's set of ESE approved benchmarking material; o a Scoring Content Specialist or Scoring Supervisor has approved using the response in a Recalibration Set. **Administration Procedure** (Performed on a per item basis.) • Before scoring commences each morning, Scorers who receive at least four (4) of five (5) exact scores on the Recalibration Set will be given access to operational images and may begin scoring for the day. • If Scorers fail to receive an exact score match on at least 4 of the 5 responses, they will be retrained by discussing the Recalibration responses in terms of the score point descriptions and the original Anchor Set. Scoring leadership will determine whether or when they may be given access to operational images to begin scoring. • Scoring leadership will carefully monitor Scorers who did not pass a Recalibration Assessment and will increase the number of read-behind responses for those individuals. • Recalibration sets are only used for operational and equating items. Field test items, with a smaller number of student responses, typically are scored entirely on the day they are trained, negating the need for recalibration. • Recalibration sets are not used for PARCC license items appearing on the MCAS test. These items will use embedded validity papers consistent with PARCC scoring For Grade 3-8 ELA and Math PARCC embedded items, Pearson will use a Validity Set approach in place of Recalibration sets and Compilation Reports. Details are provided in Appendix J. Validity Papers may also be used on Next generation MCAS items for Operational Scoring, in place of Recalibration sets, at the request of the ESE.
Read-Behind Scoring	• Read-behind scoring will allow the STLs and Scoring Supervisors to monitor each Scorer's scoring performance with an immediate real-time snapshot of the Scorer's accuracy and offer opportunities to provide individualized scoring consultations with errant Scorers.

	• If there is exact agreement between the Scorer and the STL/Scoring Supervisor score, no action is taken—the Scorer's score remains. • If there is a difference in scores, either adjacent or discrepant, the STL/Scoring Supervisor score becomes the score of record. • If the scores are discrepant, or if there are a significant number of adjacent scores (scores within one score point) between the Scorer and the STL/Scoring Supervisor , the STL/Scoring Supervisor will discuss the rationale with the Scorer. • The number of read-behinds for each Scorer will vary depending on the accuracy of the Scorer. The average per accurate Scorer is usually 10 read-behinds per full scoring day. • The minimum number of read-behinds (10 per item in a full scoring day) will be conducted for consistently accurate Scorers. When Scorer accuracy rates fall below the established standard (70% exact/90% adjacent agreement), scoring leadership increases the number of read-behinds conducted on a Scorer. If the Scorer continues to have problems maintaining accuracy on read behinds after several retrainings, the Scorer will be dismissed from scoring that item. • If Scorers meet or exceed that standard, they will continue receiving operational images on their screen and they may continue to score. • If Scorers fall below the standard threshold, they will be retrained. Scoring leadership will determine when they may be given access to resume operational scoring. In addition, if Scorers fall consistently below the accuracy standard, scoring leadership has the ability to void their scores for the day. Any voided work is processed by the scoring system as unscored responses, and are routed to scorers in the same random distribution as other unscored responses. The re-scoring of the work is conducted over the course of the remaining scoring window • Scoring Content Specialists monitor STL/Scoring Supervisor scoring accuracy and consistency by reviewing the read-behind documents they have already scored or by reading behind the STL/Scoring Supervisor whenever possible. • For Grades 3-8 ELA and Math PARCC items scored by Pearson, the Validity Paper Approach (Appendix J) will serve as the primary method for determining scorer accuracy and determine when scorers need to have their work voided. This approach would also be used for MCAS items in Grades 3-8 ELA if the ESE elects to use Validity Papers. Read behinds and Recalibration sets will continue to be the primary monitoring tool for Science and all Grade 10 assessments.
Double-Blind Scoring	• Double-blind scoring will refer to the method of scoring whereby two Scorers score the same response without any knowledge of the other Scorer's score. • In grades 3-8, 10% of responses will be double blind scored. • In High school, 100% of the responses will be double blind scored. • If there is a difference in Scorer scores that is greater than one, then the response will go into an Arbitration queue. The STL/Scoring Supervisor will score all Arbitration.

	• The percentage of responses going to Arbitration as a result of a difference of actual scores (not blanks, unreadables, etc.) should not exceed ten-percent (10%). If a Scorer's Arbitration percentage exceeds this, scoring leadership will counsel, retrain, or dismiss the Scorer. • The Resolution Chart ***(****Appendix B****)*** describes how adjacent and discrepant scores are processed. • If Scorers meet or exceed the standard, they will continue receiving operational images on their screen and they may continue to score. • If Scorers fall below the standard, they will be retrained. Scoring leadership will determine when they may resume operational scoring. • For computer based tests in grades 3-8 ELA and Math, the ePEN scoring system is able to automatically assign blank condition codes to responses that were not attempted by students. For verification, 10% of auto-scored blank responses will be double scored by human scorers.
Embedded CRRs and Seeded Papers	• Ten (10) embedded responses will be distributed at random points throughout the first full day of scoring to ensure that Scorers are sufficiently calibrated at the beginning of the scoring period. • If a Scorer falls below the required scoring accuracy rate, scoring leadership will stop the Scorer from scoring and counsel them on their scoring before allowing them to continue scoring responses. Once Scorers are allowed to resume scoring, scoring leadership carefully monitors them by increasing the number of read-behinds. All of this information will be documented and reported to the Agency. • For Equating Constructed Response 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. In ELA and Math (Grades 3-8), seeded papers will not be used in 2016-17 since no equating will be done from open response items.
Compilation Report	• Another monitoring tool will be a daily and cumulative Compilation Report that combines a Scorer's Recalibration Set results (see Recalibration Sets) with their read-behind performance. • Once Scorers have scored at least five (5) Recalibration Set responses at the start of each shift (day or evening), their subsequent read-behind scores will automatically be added to their Recalibration Set scores, and the Compilation Report will be generated. • The compilation-passing standard will be 70% exact/90% adjacent agreement (except for 1-pt math items). • If Scorers meet or exceed that standard, they will continue receiving operational images on their screen and may continue to score. • If, at the point they hit the minimum number of read-behinds, they fall below the 70% exact/90% adjacent-agreement threshold, scoring leadership will stop them from scoring and tell the Scorers they need to see the Scoring Supervisor for possible retraining. • The Scoring Content Specialist/scoring manager/scoring director may approve, on Scoring Supervisor recommendation, the resumption of scoring for Scorers who fell below standard based on Compilation Report

	data. • Scoring leadership will carefully monitor each individual whose scoring was stopped due to below-standard scoring and will increase the number of read-behind responses for these Scorers. • A final Compilation Report will be run at the end of each day. If there are individuals who are still below the 70% exact/90% adjacent agreement level, scoring leadership will void their scores for that specific shift and return the responses to the scoring queue for re-scoring by other Scorers on subsequent scoring days. • The compilation report will not be used for items in Grades 3-8 that use the Validity Paper approach for quality control.
Scorer Monitoring Reports	• Because the MCAS is a complex testing program, computer-generated reports will be necessary to ensure that o overall accuracy, consistency, and reliability of scoring (group level) is maintained at an acceptable level, o immediate, real-time Scorer data (individual level) is available to allow early Scorer intervention when necessary, and o Scoring schedules are maintained. • Some reports will be available to STLs and Scoring Supervisors at the scoring tables; others will only be available to Scoring Content Specialists, Scoring Managers, and Scoring Director. • The following types of reports are available to Scoring staff: o The Read Behind Summary *(Appendix D)* report shows the total number of Read Behind responses read by both the Scorer and the STL/Scoring Supervisor and will note the number and percentage of exact, adjacent, discrepant scores. o The Double Blind Summary *(Appendix E)* report shows the total number of Double Blind responses read by a Scorer and will note the number and percentage of exact, adjacent, and discrepant scores. o The Compilation Report *(Appendix F)*will show for each Scorer, the total number of responses scored, the number of read-behind responses, the number of scored Recalibration assessment responses (at least 5 per day), the total number of read-behind and Recalibration assessment responses,, the percentage of exact, adjacent, and discrepant scores. o The Embedded CRR Summary *(Appendix G)*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. o The Qualification Statistics Report *(Appendix H)* will list each Scorer by name and ID#, identify which Qualifying Sets they did or did not take, and for the ones they did take, whether they passed or failed. o Summary Report *(Appendix I)* will list the total number of student responses loaded into the scoring system, the number of responses scored, and the number of responses to be scored.

- Pearson scorer Compilation report (Appendix K) will list the combined scorer performance on Validity Papers, Read Behinds, and Double Blind Scoring

Scoring Facilities

- Measured Progress and Pearson will be responsible for identifying the scoring facilities and making all Scorer and facility arrangements for the grades and content areas scored
- The following criteria will be reviewed in identifying appropriate facilities:
  - security
  - Scorer capacity
  - proximity to qualified, diverse scoring pool
  - capacity for electronic equipment
  - access to major highways, public transportation, and parking
  - accessibility to individuals with disabilities
  - proximity and adequacy of restrooms
  - adequacy of lighting
  - adequacy of heating and air-conditioning
- Measured Progress and Pearson maintains several permanent and seasonal scoring locations.
- Constant daily communication and coordination is done through secure network e-mail, telephone, and messaging services to ensure that any information and/or modifications to scoring processes can be implemented across all scoring locations.
- Use of distributed scoring (for PARCC licensed items) involves remote Scorers complying with security measures, signing non-disclosure agreements, and use of technology tools to ensure security of items and responses.
- Planned Scoring Center locations for the following grades/contents are detailed in the chart below

<table>
<tr><th></th><th colspan="3">Operational Scoring</th><th colspan="3">Field Test Scoring</th></tr>
<tr><th>Grade</th><th>Math</th><th>ELA</th><th>Science</th><th>Math</th><th>ELA</th><th>Science</th></tr>
<tr><td>3</td><td rowspan="2">Mesa, AZ</td><td>Mesa, AZ</td><td></td><td rowspan="8">Longmont, CO</td><td rowspan="8">Longmont, CO</td><td></td></tr>
<tr><td>4</td><td>Virginia Beach, VA</td><td></td><td></td></tr>
<tr><td>5</td><td rowspan="4">San Antonio, TX</td><td rowspan="2">Columbus, OH</td><td>Menands, NY</td><td>Menands, NY</td></tr>
<tr><td>6</td><td></td><td></td></tr>
<tr><td>7</td><td rowspan="2">Charlotte, NC</td><td></td><td></td></tr>
<tr><td>8</td><td>Menands, NY</td><td>Menands, NY</td></tr>
<tr><td>10</td><td>Longmont, CO</td><td>Reading: Longmont, CO
Writing: Menands,</td><td>Menands, NY</td><td></td><td></td><td>Menands, NY</td></tr>
</table>

	NY
Scoring Security	• All scoring centers will be secure and locked with admissions limited to those employed in the scoring process. Scoring center entrances are kept secure by an electronic swipe card system, an employee posted at the door, or both. • Upon arrival at the scoring site, all Scorers must provide identification. It is only upon submittal of such identification that Scorers are provided with the requisite scoring center materials. • All Scorers must log on to the image-scoring system using an assigned username and password that is issued and carefully monitored by Measured Progress. • Utilization of the image-based scoring system will minimize the need to handle physically actual student response booklets and related scoring materials. • No print copies of items or responses leave the premises unless designated for public release or for use with the ESE's professional development program. Under those circumstances, procedures such as those for test production and off-site test review are followed. • Barrels, boxes, or other receptacles are provided at strategic locations throughout the scoring sites for all test-secure materials needing to be safely destroyed. • All print copies of items or responses will be destroyed by shredding or other means by an outside vendor who will provide proof of destruction.
Scoring Staff	• Scoring Project Manager: Overall responsibility for all scoring-related activities and deliverables for a specific contract(s) • Content Group Manager: Primarily responsible for coordinating efforts of Scoring Content Managers and Scoring Content Specialists. • iScore Operational Manager: Sets up and maintains iScore system for scoring and acts as Scoring's liaison with Technology. • Assistant Director, Scoring Content: Primarily responsible for coordinating efforts of scoring sites, schedules, processes, and procedures. • Scoring Site Managers: Oversees site logistics, facility requirements, and security. • Scoring Content Managers (Language Arts; Math and Science): Responsible for managing all staff within their assigned content area(s) and working to ensure staffing and requirements are being met. • Scoring Content Specialist (Language Arts): Responsible for maintaining scoring accuracy and reliability in language arts to meet state educational standards. Identifies and selects training materials, conducts scoring leadership and Scorer training, and monitors the scoring of Scoring Supervisors, STLs and Scorers. • Scoring Content Specialist (Math): Responsible for maintaining scoring accuracy and reliability in Math to meet state educational standards. Identifies and selects training materials, conducts scoring leadership and Scorer training, and monitors the scoring of Scoring Supervisors, STLs and Scorers.

	• Scoring Content Specialist (Science): Responsible for maintaining scoring accuracy and reliability in Science to meet state educational standards. Identifies and selects training materials, conducts scoring leadership and Scorer training, and monitors the scoring of Scoring Supervisors, STLs and Scorers. • Scoring Supervisor (Measured Progress and Scoring Directors (Pearson): Responsible for identifying and selecting training material, training and monitoring Scorers, and monitoring iScore performance for a small group of Scorers. • Scoring Team Leader (STL Measured Progress) and Scoring Supervisors (Pearson): Responsible for assisting the Scoring Supervisors with identifying and selecting training material and monitoring the small group of Scorers (no more than 11 Scorers to each STL).

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 Measured Progress is placing the Temporarily Placed Employee Measured Progress to provide services on a temporary basis; and

Whereas the Temporarily Placed Employee may, in the course of performing services for Measured Progress, have access to Measured Progress' confidential, proprietary information; and

Whereas Kelly has agreed to assure that the Temporarily Placed Employee preserves Measured Progress' confidential proprietary information; and

Whereas Kelly wishes to provide Measured Progress with assurance that Kelly is instructing its employees of the importance of preserving Measured Progress' confidential and proprietary information; and

Whereas the Temporarily Placed Employee, as an express condition of being placed to provide temporary services for Measured Progress, whether remotely or from the business premises of Measured Progress, agrees to preserve Measured Progress' 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 Measured Progress' premises, in the course of performing assigned services, may have access to Measured Progress' valuable, proprietary, and confidential business information and trade secrets as well as confidential test materials and results used by Measured Progress in servicing its clients (including information prepared by Measured Progress or supplied to Measured Progress 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 Measured Progress' guidelines and procedures regarding access to, and use of, Measured Progress'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 Measured Progress. Kelly and the Temporarily Placed Employee understand and agree that any and all Proprietary Information is the property of Measured Progress. 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 Measured Progress. 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 Measured Progress.

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 Measured Progress promptly so that the it may seek a protective order or other appropriate remedy or, in Measured Progress'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 Measured Progress 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 or 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 Measured Progress and prior to his or her departure, the Temporarily Placed Employee agrees to return to Measured Progress all copies (in whatever form, including electronic media) of any Proprietary Information to which the Temporarily Placed Employee had access while performing services for Measured Progress.

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 Measured Progress; (iii) that is disclosed with the express written permission of Measured Progress; 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 Measured Progress 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 Measured Progress, 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 Measured Progress its reasonable attorney fees incurred to obtain such injunctive relief.

9. In addition, the parties recognize that Measured Progress 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 Measured Progress'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 Measured Progress is an intended third party beneficiary of this Agreement and has the right to enforce its provisions.

Agreement dated as of _____,

Kelly Services

By:_____

Title:

Temporarily Placed Employee

_____

Appendix B

MCAS Resolution Charts

Read-Behind Scoring [1]		
Scorer Score	**STL/Supervisor Score**	**Final**
4	**4**	**4**
4	**3**	**3**
4	**2**	**2**
[1]In all cases, the STL score is the final score of record. If there are conflicting resolution scores, it will be arbitrated by senior leadership.		

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.			

Writing English Conventions Double Blind Scoring (High School Legacy Test)			
Scorer #1	**Scorer #2**	**Scoring Supervisor/STL Resolution**	**Final**
4	**4**	**-**	**8**
4	**3**	**-**	**7**
4	**2**	**4**	**8**
4	**2**	**3**	**7**
4	**1**	**3**	**7**
4	**1**	**2**	**3**
[1]Identical or adjacent Scorer scores are summed to obtain the final score. The resolution score, if needed, if summed with an identical Scorer score; or, if the resolution score is adjacent to one or both Scorers but not identical, than it is summed with the highest adjacent score to reach the final score			

	Writing Topic Development Blind Scoring (High School Legacy Test)			
Scorer #1	**Scorer #2**	**Scoring Supervisor/STL Resolution**	**Scoring Content Specialist Score**	**Final**
6	**6**	**-**	**-**	**12**
6	**5**	**-**	**-**	**11**
6	**4**	**4**	**-**	**8**
6	**4**	**5**	**-**	**11**
6	**2**	**4**	**4**	**8**
6	**2**	**4**	**3**	**6**
6	**2**	**3**	**-**	**5**
	[1]Identical or adjacent Scorer scores are summed to obtain the final score. The resolution score, if needed, if summed with an identical Scorer score; or if the resolution score is adjacent to Scorer #1 and/or #2 but not identical to either,			

	then the two highest adjacent scores are summed for the final score. If the resolution score is still discrepant to both scores, the Scoring Content Specialist assigns a 4th score, which is doubled to obtain the final score.

Appendix C

Final Score Determination for [Project] Responses

#	Available Scores for a Response	Scores	Final Score iScore
1	**Edit score**	**Scoring Content Specialist/ Edit Scoring Supervisor**	**Edit score is provided**
2	**Arbitration score** (No Edit, Wrong Location or Read Behind Scores)	**Scoring Team Leader (Scoring Supervisor)**	**Arbitration score is provided**
3	**Arbitration score** **Read Behind score(s)** (No Edit or Wrong Location score)	**Scoring Team Leader (Scoring Supervisor)**	**Read Behind score(s) is supplied when available** **Arbitration score(s) is supplied when no Read Behind score(s) is available**
4	**Read Behind score(s)** (No Edit, Wrong Location or Arbitration score)	**Scoring Team Leader (Scoring Supervisor)**	**Read Behind score is provided**
5	**Primary Scorer score** **Double Blind Scorer score** (No Edit, Wrong Location, Arbitration or Read Behind scores)	**Scorers**	**Primary and Double Blind scores are provided**

Appendix D

Read Behind Summary Contract: 1705 Spring Grade: 07
Content: Language Arts (Example)

Name	ID#	Total Scored	Read Behind							
			Total DB	% DB	Exact	% Exact	Adj	% Adj	Disc	% Disc

Appendix E

Double Blind Summary Contract: 1705 Spring Grade: 07
Content: Language Arts (Example)

Name	ID#	Total Scored	Double Blind							
			Total DB	% DB	Exact	% Exact	Adj	% Adj	Disc	% Disc

Appendix F

Compilation Report Contract: 1705 Spring Grade: 07
Content: Language Arts (Example)

MPID	Total Scored	Read Behind	Recalibration	Total Compilation	Exact	pExact	Adj	pAdj	Disc	Action

Appendix G

Scorer Embedded CRR Statistics Contract: 1705 Spring Grade: 07
Content: Language Arts (Example)

Name	ID#	Total Scored	Emb CRR							
				%		%		%		%
			Total CRR	CR	Exact	Exact	Adj	Adj	Disc	Disc

Appendix H

Scorer/Item Qualification Summary Contract: 1705 Spring
Grade: 07 Content: Language Arts (Example)

ID#	W R 1 1 0 0 1	Total Completed CRRs	# Pass	# Fail	% Pass
Total Passed	0	0	0	0	0
Total Failed	0	0	0	0	0
Total Number	0	0	0	0	0

Appendix I

Response Load Status Report Contract: 1705 Spring
Grade: 07 Content: Language Arts (Example)

ResponseCode	RemainInRespQ	ScoredinRespQ	RemaininDBQ	ScoredInDBQ	Updated
Total Responses					
Total Responses %					

Appendix J – Pearson Scoring Guide

Condition Codes

The following hierarchy for adjacency is: UN<< OT<< NE<< DC<< BL<< 0< 1< 2< 3< 4< 5

Condition code labels and values are UN, OT, NE, DC, and BL.

BL – Blank = No attempt to respond to item. Either nothing or stray marks appear on paper.

DC – Direct Copy = Direct copy of test from the passage or question.

NE – Non-English = Response written entirely in a language other than English, or without enough English/numbers to provide a score.

OT – Off-Topic = Response does not address the topic or task for item. Response is irrelevant to the prompt. Response states student is refusing to participate in testing.

UN – Unreadable = Response cannot be read because of poor penmanship, spelling that cannot be deciphered, writing is too small, too faint to see, or only partially visible.

Scorers will only have access to score condition code BL (Blank). All other condition code responses will be sent to the review queue in ePEN and scored by Supervisors and Scoring Directors.

ELA - If one trait in a multi-trait item receives a condition code, it applies to all traits.

Math – If one trait in a multi-trait item (or a composite item with multiple OE components) receives a condition code, the condition code does NOT apply to the all traits.

PARCC and MCAS Single vs Multi-trait items

PARCC Math items are made up of single trait and composite items.

PARCC ELA items are multi-trait with 2 traits.

MCAS Math items are all single trait.

MCAS ELA items are both single and multi-trait.

PARCC Item Notes

PARCC scoring consists of ELA and Math.

ELA: Grades 3 through 8

ELA items are scored depending on the task type.

Narrative (two traits)

Expression trait has a 0,1,2,3 score scale for grades 3-5 and a 0,1,2,3,4 score scale for grades 6-8

Conventions trait has a 0,1,2,3 score scale for all grades

Math: Grades 3 through 8

Math items are scored based on one or more scoring guides. Each item has its own specific rubric. Math items are either single-trait (non-composite) or multi-trait (composite). The rubric will define whether an item is single or multi-trait. Math scoring guides for each trait are as follows:

2-category has a 0,1 score scale
3-category has a 0,1,2 score scale.
4-category has a 0,1,2,3 score scale.
5-category has a 0,1,2,3,4 score scale

PARCC will utilize ePEN2 and PLS for scorer training, and ePEN2 for scoring.

PARCC will train utilizing baseline items in ePEN2. Baseline items are also referred to as prototype items. A prototype item will have one or more items linked to it within the ePEN2 system for purposes of qualification.

There are two different types of training sets – prototype and abbreviated. Items with prototype training sets are linked to specific items with abbreviated training sets. Only prototype items have qualification sets; therefore, a scorer must qualify on a prototype item in order to score that prototype or any abbreviated item associated with that prototype.

PARCC Scorer Qualification Standards

Scorers in all content areas must meet the following qualification requirements prior to scoring:

ELA– must achieve at least 70% perfect agreement on two of three qualification sets and 96% perfect plus adjacent agreement across all scores assigned for three qualification sets. A 70% perfect agreement must be achieved on each trait at least once across the three sets.

Math– must achieve the following exact and exact plus adjacent agreement percentages after completing three qualification sets:

2-category (0-1 scale): 90% exact; 96% exact + adjacent
3-category (0-2 scale): 80% exact; 96% exact + adjacent
4-category (0-3 scale): 70% exact; 96% exact + adjacent
5-category (0-4 scale): 70% exact; 95% exact + adjacent

Exact agreement must be achieved on two of the three qualifying sets for each category score point scale associated with the prototype item.

Exact + adjacent agreement is not measured at the set level, but rather an average of all three qualification sets. To qualify, a scorer must be adjacent to the true score on 29 of the 30 qualification responses.

Math exact and exact + adjacent requirements must be met for each Part (trait) associated with the item. A single trait item will only have one Part to calculate. A multi-trait item may have up to four Parts to calculate.

For example, if an item has two Parts where Part A is a 2-category scale and Part B is a 4-category scale then the scorer must achieve 90% exact agreement on two of the three sets for Part A, and 70% exact agreement on two of the three sets for Part B. In addition, a scorer must be adjacent on 29 of the 30 qualification responses for Part A and Part B.

Scoring Rules for both PARCC and MCAS items

Auto-scoring of blanks will be enabled for online items. ePEN randomly selects 10% of the responses per item for 2nd scoring. Human scored blanks with a 1st score as blank are also eligible to be selected for 2nd scoring.

Pearson will conduct double-blind scoring on 10% of all 3-8 Math and ELA items.

Resolution scoring of all Pearson scored items is required:

If the 1st score is non-adjacent to the 2nd score, the response is routed to the Resolution Queue for a blind 3rd read. For multi-trait items, when any one or more traits have a non-adjacent 1st or 2nd score, the response is also routed to the Resolution Queue for a blind 3rd read. If one trait receives non-adjacent scores, the entire item goes to resolution, not just the trait that was non-adjacent. Therefore, resolution of these multi-trait items is by item, not by individual trait.

Final scoring for all items:

If there is a Backread (Read Behind) score, it is the final score
Else if there is a Resolution score (Arbitration), it is the final score
Else if there is a 1st and 2nd score, the highest score is the final score
Else the first score is the final score

PARCC items are scored in a distributed environment and online training is being utilized.

MCAS items are scored in regional sites and stand-up training is utilized.

Pearson's Scorer to Supervisor ratio will not exceed 20 to 1.

Read-Behind Scoring

Read-behinds are conducted for all individuals scoring 3-8 Math and ELA items. Pearson targets 5% daily read-behinds for every scorer and a minimum of 5 responses per scorer per day. Read-behinds are a tool used by leadership to monitor scorer accuracy to detect the need for additional training. Increased read behinds are directed towards score's whose validity perfect agreement rate drops to within 5% of the perfect agreement standard. Supervisors will perform additional oversight of those scorers to ensure project quality is not compromised.

Validity

Pearson will embed validity papers on all items scored. The following quality metrics 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 / Math Multi Trait	2- traits	65%	65%	96%

Validity Insertion and Selection for MCAS items

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.

Validity Insertion and Selection for PARCC items

PARCC items have embedded validity at a rate of 4% for ELA and 4% for Math. The validity papers for these items are rolled over from the PARCC administration along with the training materials. As Massachusetts specific scoring trends arise, Supervisors will escalate appropriate responses 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, it is moved to the "Pending" status and the Content Specialist will "Approve" or decline its use. At any time, the Customer always has ePEN access to review these approved validity papers.

Validity Intervention Standards

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.

Appendix K – Pearson Scorer Compilation Report

Daily Scorer Compilation

09/04/14

ItemName	EmployeeID	Score Count	Validity Count	Validity Agree %	Validity Adj %	Validity Agree + Adj	ValDisc	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%			