

3.0 Scoring and Reporting

3.1 Scoring

The role of scorers is to judge whether the evidence submitted for each mastery objective, the artifact, demonstrates that the student has attained the conditions required for mastery of that objective. The following sections outline the procedures implemented by Pearson Educational Measurement's (PEM) Performance Scoring Center (PSC) to ensure the reliability and accuracy of the scoring process and results.

Recruitment of Scorers and Scoring Supervisors

Highly qualified scorers are essential to achieving and maintaining a high degree of consistency and reliability in scoring students' responses. The careful selection of professional scorers to evaluate student portfolios is therefore essential to scoring the ALT-MSA. In the selection of candidates for scoring the Alternate Maryland School Assessment (ALT-MSA), priority is given to individuals with degrees in special education in addition to those with previous experience in performance scoring. At a minimum, all scorers have a four-year college degree and must complete the formal application process including an interview. Such prescreening of candidates ensures selection of only the highest caliber of scorers. Regardless of previous experience or education, however, all selected scorers are required to meet the project's qualification standards (acceptable scores on qualifying set) and are subject to continual monitoring (i.e., backreading and validity) for quality and accuracy.

Scoring supervisors are chosen from the larger pool of scorers based on demonstrated expertise with the ALT-MSA scoring process, organizational abilities and training skills. Individuals chosen to perform these assignments possess leadership abilities and positive interpersonal communication skills. Supervisors also possess the essential capability of helping scorers to understand the particular scoring requirements of the ALT-MSA. A list of all those involved in the ALT-MSA scoring effort and their roles is provided in Appendix E.

Recruitment for the ALT-MSA begins approximately six weeks before the onset of scorer training.

Rangefinding

Rangefinding is the process by which a wide range of portfolios are reviewed by a committee of experts for the purpose of selecting exemplars to use in the training, monitoring, and qualification of scorers and for establishing/revising the scoring guidelines. For the ALT-MSA a sample of approximately 120 portfolios are chosen by MSDE for rangefinding:

- 50 portfolios from grades 3, 4, and 5
- 50 from grades 6, 7, 8
- 20 from grade 10

To the extent possible, these portfolios represent the range of abilities and characteristics in the population tested as well as a range of artifact types. The goal is to provide the

range-finding committee with a sample of portfolios that is diverse enough to highlight any issues that may be encountered during scoring and therefore should be addressed in training. The range-finding portfolio selection process for the current administration is outlined in Appendix F.

Prior to the range-finding meeting, participating PSC staff members familiarize themselves with the range-finding portfolios, and review the training materials and scoring decisions from the previous year's scoring. They then meet with the MSDE to further review and discuss these portfolios and plan the order of portfolio presentation. The range-finding agenda is finalized at this time.

At the start of the range-finding meeting, the committee members, in conjunction with the MSDE and the PSC staff, begin work by reviewing the scoring rules and decisions from the previous year. This helps to ensure a common understanding of standards and promote consistency of scoring from year to year. Next, the range-finding committee is introduced to their tasks: 1) reviewing and scoring the range-finding portfolios to be used in the training of scorers, and 2) determining the scoring guidelines.

Throughout the meeting, PSC staff members maintain notes and records consensus scores, teacher comments, and discussions of portfolios. Teacher comments and discussion are used by staff to aid in scorer training. At the end of each day MSDE and PSC staff members debrief to discuss the committee work and any scoring issues from the day. In addition, the agenda for the next day is discussed and adjusted as needed.

At the end of the range-finding meeting PEM provides the MSDE with the official range-finding record, which includes consensus scores and teacher's comments. Both the MSDE and a PEM staff member sign this record to certify that the scores have been recorded accurately. The PEM Scoring Director will later add information on the placement of each portfolio in the training and qualifying sets.

Immediately following the range-finding meeting, the MSDE and the PSC conduct a post-range-finding session to prepare the scoring guide, training sets (i.e., anchor sets and practice sets), qualifying sets, and a validity set. The scoring guide, training sets, and qualifying sets are submitted to MSDE for approval and sign off before scoring supervisor training begins.

Training

Training begins with the distribution and review of the Scorer Participant Guide. The Scorer Participant Guide introduces potential scorers to the schedule, provides an overview of the training and scoring process, explains general PSC training, scoring and quality-control procedures, and gives specific information about Pearson Educational Measurement and the Alternate Maryland School Assessment.

The Training Process

Scorers are trained to score all grade levels in both reading and mathematics content areas. The ALT-MSA scoring rules are presented in context with student portfolios. First, an anchor set of portfolios, consisting of all training issues, is introduced to scorers.

Then, a set of practice portfolios is used to give the scorers the opportunity to practice scoring. Finally, a set of qualifying portfolios is administered to the scorers to determine if they have fully grasped the scoring criteria and rules.

Introduction

During the introduction, hard copies of all training sets are provided to the scorers for review and discussion. Scorers are encouraged to take notes throughout the training process. Scorers are also provided with

- an overview of relevant vocabulary specific to special education and the alternate assessment;
- an introduction to the Maryland State Content Standards in both reading and mathematics and an explanation as to how these standards guide the assessed objectives;
- an explanation of portfolio contents and organization;
- the criteria for acceptable evidence of mastery;
- an in-depth review and discussion of the scoring rules and guidelines; and
- an overview of the list of positive practices.

Anchor Portfolio Set and Scoring Guide

After the general introduction, the scoring director introduces the anchor portfolios in conjunction with the content standards and scoring rules. The Anchor Set is a combination of portfolios that are exemplary and portfolios with common scoring issues. Each anchor portfolio demonstrates a clear, straightforward presentation of mastery or non-mastery of the objectives. The Scoring Director discusses the uniqueness of each portfolio, highlighting critical information that demonstrates exactly why an objective is considered mastered or not. Eight anchor portfolios train scorers to understand the criteria for scoring and provide references for use during live scoring.

Practice Portfolio Sets

As part of training, scorers practice scoring on sets of practice portfolios. Through two practice sets of five portfolios each, scorers hone their skills to understand the scoring guidelines, content standards, and evidence of mastery. Scorers score the practice sets independently using the anchor set, the content standards, and the scoring rules as guidelines. Scoring the practice portfolios is not as clear as the anchor portfolios. Practice portfolios contain questionable objectives and artifacts that may not be straightforward. During practice, questions and interaction are encouraged so scorers may further internalize the scoring guidelines. The Scoring Director reviews the scorers' practice portfolios and provides the correct scores. Practice is an essential part of the training procedure.

Qualifying Portfolio Sets

After practice and review, scorers take a qualifying set of three portfolios. Again independently, the scorer uses all training materials to score the qualifying set. Each qualifying set consists of three complete portfolios. For a scorer to begin live scoring 80% perfect agreement is required on one of two qualifying sets. After each qualifying set, a review of the scores takes place in order for scorers to understand their errors. If a

scorer does not qualify on the first set, the scoring director reviews that scorer's errors with him/her before administering a second qualifying set of three portfolios. Scorers not meeting the established guidelines by the end of the training session are dismissed. The percentage of scorers that qualified to score the current administration and the average qualification score (i.e., percent agreement) is provided in Appendix A, Table 2.

Once scorers have qualified, the scoring director trains the portfolio flow, including how to first and then second score and the alert process. Scorers are then divided into teams based on performance on the qualifying sets, prior experience, and specific areas of expertise. Teams are constructed to be as similar as possible given these variables. A scoring supervisor is assigned to each team and, at this point, scorers begin live scoring.

Training of Scoring Supervisors

Scoring supervisors receive the same content and scoring training as scorers, in addition to extra training on supervisory duties. Each supervisor receives extensive training on the material circulation. A select group of scoring supervisors also receives additional training on resolution scoring.

Distribution of Portfolios to Scoring Teams

Upon arrival at the scoring site material handlers unload and check in student portfolios. Boxes arrive in numbered batches. Material handlers check each portfolio in on a shipping list and then file it in a secure warehouse according to batch number until scoring.

At scoring time, material handlers deliver a batch of approximately 24 portfolios to the scoring supervisor of a team. The supervisor signs off receipt of the batch on the Warehouse Batch Tracking Log. Scorers sign out an individual portfolio on a Batch Tracking Log that remains with each batch. They then return completed portfolios to an area designated "first score complete." Material handlers collect the portfolios and return them to the warehouse to be refiled. When all of the portfolios associated with a batch have been returned to the warehouse, the batch is delivered to a different scoring team for second scoring. No team reviews the same batch of portfolios twice.

For the 2003-2004 administration, grade and order of receipt determined the order in which portfolios were scored. A sample of the portfolios from grades 3, 5, 8, and 11 was required to accommodate standard setting, therefore scoring priority was given to these grades. Upon receipt and check-in at the scoring site, portfolios from these grades were immediately batched and distributed for scoring.

Scoring Procedure

The ALT-MSA Scoring Process is defined in Appendix G. This document chronologically defines the steps a reader should follow to review a portfolio and score the associated artifacts. It also delineates the scoring rubric and provides examples of mastery objectives/artifacts that would receive a condition code rather than a score.

Each artifact within a portfolio is scored at least two times. Portfolio artifacts for which the first and second scores do not agree are sent to resolution. Resolution readings are

identified by the supervisors and performed by the Scoring Director, Assistant Scoring Director, PSC Project Manager, Scoring Supervisors, or designated agent (experienced scorers). The Scoring Director supervises all individuals performing resolution readings.

Some mastery objectives may not be scorable according to MSDE criteria. If a scorer believes that a mastery objective is not scorable, for whatever reason (i.e., alignment issues, artifact not dated or name missing, or as determined by current administration scoring rules), the scorer brings the portfolio to his/her supervisor for review. If the supervisor is uncertain how to score the objective, the Scoring Director is consulted. If a score or condition code cannot be determined based on established scoring rules, the MSDE is consulted. Any scoring decisions or policy rulings are documented by the Scoring Director.

After the appropriate score or condition code is determined by supervisory staff, the score or code is recorded on both the first and second scoring monitor by the scoring supervisor. (The scoring monitor is the scannable document that allows each student's scores to be captured electronically.) This ensures that a second scorer will not be bringing the same issue to the attention of supervisors and the Scoring Director after it has already been reviewed by supervisory staff.

Quality Control

Backreading

Backreading is a source of information on scoring accuracy. Backreading is one of several methods used to ensure reader accuracy whereby a scoring supervisor reviews a random sampling of scores assigned by readers on their team to assess accuracy. Backreading is trained during scoring supervisor training, is initiated at the beginning of scoring, and continues throughout scoring. It is a PEM standardized ISO procedure used to monitor scorers, to help eliminate drift by alerting scorers to their mistakes at the team level, and anchoring them back to the training materials and scoring rules. Backreading results are documented and recorded by supervisors on backreading tally forms.

Each day every team reviews the training sets and scoring rules. Reviewing the training materials keeps all scorers and scoring supervisors grounded in the guidelines established during training. If a scorer is absent for two days or more, he/she reviews all training materials and scoring rules with a supervisor, updating the scorer on any missed scoring decisions. The scorer also takes a validity portfolio to ensure he/she is still scoring accurately.

Validity Sets

Validity portfolios are portfolios whose "true scores" have already been determined by the Scoring Director and the MSDE. These portfolios are interspersed among the portfolios to be scored to allow individual scorer accuracy to be assessed throughout the scoring process. The average percent agreement between readers' scores and the "true scores" for these validity sets is provided in Table 3 of Appendix A for the current administration.

Validity reports and other reports generated by the Electronic Paper Scoring System (ePS) are described below.

Reports Generated and Used by PSC Staff to Monitor Scorers and Scoring Accuracy and Control Scorer Drift

In 2003 - 2004 the PSC staff reviewed and distributed reports daily to evaluate reliability and other scorer statistics. However, the length and complexity of the reports made it difficult to determine overarching issues on the project. These reports were originally designed for a regular assessment and proved to be inefficient for the ALT-MSA. As a result, the scoring performance reports for future years are being enhanced to provide summary information at the portfolio and mastery objective level. These enhanced summary reports will also provide team statistics so that these can be compared to the scoring room, as a whole. These will allow MSDE and the PSC to effectively work together to determine scoring issues and reduce the number of resolutions.

- Score reliability reports:

The inter-rater reliability reports document how often two scorers agree when scoring the same response.

The Scoring Director reviews reliability reports daily to ensure that all items being scored are within the acceptable reliability parameters. If an item's reliability falls significantly below an acceptable level, an action plan is created. If a scorer's reliability falls significantly below the room average, the Scoring Director reanchors them using the relevant training materials and scoring rules established during the initial training.

- Frequency distribution reports:

The frequency distribution reports document the percentage of scores given that falls into each score point and condition code category.

The Scoring Director reviews the frequency distribution reports. If a scorer is scoring too high or too low compared to the rest of the group, retraining may occur.

- Validity reports:

Validity reports can be a useful monitoring tool. Validity reports document how often a scorer agrees with the "true score" given to preselected validity responses.

The Scoring Director reviews the validity reports to identify struggling scorers and determine whether there is any room drift or a particular type of item or issue causing problems. A struggling scorer is a scorer below the validity requirement and/or significantly below the room average. When identified, the Scoring Director and scoring supervisors monitor and provide remediation to struggling scorers. Room drift occurs when a group of scorers consistently score an

objective or item incorrectly on validity. If there is strong evidence of room drift, project management may consider retraining or calibration of that particular objective or item.

All reports are monitored by the Scoring Director and Project Managers throughout the scoring process. The reports are also discussed with the MSDE on a regular, ongoing basis. Based on these reports, backreading, and trends found in resolution scoring, it may be necessary to retrain on a particular item or create a calibration set. If needed, calibration sets are created by PSC staff and approved by MSDE staff. Calibration is a form of training that creates consensus and accuracy within the scoring pool (both scorers and supervisors). A calibration set focuses on one problem or issue. Calibration papers or portfolios are focused with a single, clear purpose. A list of the steps taken by the PSC to ensure scorer accuracy and correct for scoring drift is provided in Appendix H.

Security at the Scoring Site

Providing an environment that promotes the security of test items, student responses, data, and employees is of utmost concern to PEM. We employ the following standard safeguards for security at all of our sites:

- Controlled access to the facility.
- Materials leave the facility during the project only with the permission of the customer (Maryland State Department of Education).
- Scoring personnel sign a nondisclosure and confidentiality form in which they agree not to use or divulge any information concerning tests, scoring guides, or individual student responses.
- All staff display PEM identification badges at all times while in the scoring facility.
- No recording or photographic equipment is allowed in the scoring area without the consent of the customer (MSDE).
- No cell phones of any kind are allowed in the scoring area.
- All contact with the press is handled through the customer (MSDE).

3.2 Standard Setting

Proficiency levels were established for the Independence Mastery Assessment Program (IMAP) in Summer of 2003. IMAP was the predecessor assessment to the ALT-MSA. This process involved Maryland educators applying a portfolio paper sorting method to the 2002-2003 assessment results. In order to ensure uniform performance standards between IMAP and ALT-MSA, a process of linear transformation was used to translate the IMAP growth score proficiency level cut points to the ALT-MSA mastery percentage proficiency level cut points. This process resulted in two performance standards on the mastery percentage scale that define the basic, proficient, and advanced proficiency levels described below.

Basic: *Students at this level demonstrate 0% to 50% mastery of the skills tested in reading and mathematics.*

Proficient: Students at this level demonstrate 60% to 80% mastery of the skills tested in reading and mathematics.

Advanced: Students at this level demonstrate 90% or greater mastery of the skills tested in reading and mathematics.

3.3 Reports

A variety of reports are described and listed in this section. Samples of some of these reports can be found in Appendix I of this document.

Description and Interpretation of Scores

The following scores are calculated and reported to students, schools, and/or districts that participate in the ALT-MSA.

Mastery Objective Score

Each student who participates in the ALT-MSA is assessed on 20 unique mastery objectives: 10 for each subject area. A mastery objective is a clear statement of the specific response a student must provide (and the conditions under which it must be provided) in order to demonstrate mastery of a particular objective. For each mastery objective assessed, an appropriate artifact is submitted in the student's ALT-MSA portfolio for scoring. The artifact is scored as either exhibiting mastery or non-mastery of the associated objective. If mastery status cannot be determined the student is assigned a not-scorable condition code for that mastery objective (see Appendix G).

Students must select two objectives from each subject area for demonstration in one additional setting other than the classroom. For the selected objectives, students submit two artifacts; one corresponding to each setting in which evidence was collected. In order for a student to achieve a score of mastery on a "multiple setting" objective, *both* of the artifacts submitted for that objective must be scored as "mastered." If one of the artifacts is scored as "not mastered," a score of 0 (not mastered) is assigned to that objective.

By themselves mastery objective scores provide only an indication of whether or not the artifact submitted for a given mastery objective met the requirements for mastery. Unless a condition code is provided, no further information can be gleaned from this score. Specific information regarding how and why mastery was (or was not) obtained must be determined from the submitted artifact and its level of accuracy (i.e., the value compared to the 80% mastery criterion).

Given the purpose of the ALT-MSA, and therefore the manner in which mastery objectives are developed and assessed, one must be careful not to generalize mastery objective scores beyond the specifics of the task assessed. Although mastery objectives are developed to map back to the Maryland State Content Standards, success on a specific mastery objective may not generalize to a similar task measuring the same underlying objective. In order to make generalizations regarding a student's knowledge and skills with respect to an underlying objective further evidence of success is typically required.

Average mastery objective scores for the current administration can be found in Appendix A, Tables 4 and 5 for Reading and Mathematics, respectively. These averages are provided by content standard and therefore do not include students assessed on access skills. For each content standard the value provided indicates the percentage of all artifacts associated with that content standard that were scored as “mastered.” For example, if the average mastery objective score associated with the Phonics/Phonemic Awareness standard were 0.85, this would indicate that 85% of the submitted mastery objectives associated with this content standard were scored “mastered.”

Mastery Percentage Score

Within each subject area the proportion of mastery objectives scored as “mastered” (i.e., that have an artifact that meets the criteria outlined for mastery) is the mastery percentage score for that subject. Mastery percentage scores are used to categorize students into one of three different proficiency levels: Basic, Proficient, and Advanced. Each proficiency level identifies a particular range of mastery percentage scores that corresponds to a level of academic achievement. (See section 3.2 of this document for a description of standard-setting process and the resulting proficiency level definitions.) The ultimate goal of NCLB is for all students to reach the Proficient or Advanced level.

The ALT-MSA is intended to assess each student on a set of skills and objectives that are appropriate, yet challenging. As a result, the specific set of mastery objectives assessed is different for each student. This would seem to suggest that a given student’s mastery percentage should not be compared to that of another student or the state/system/school average. To an extent this is true. It is quite possible that the set of mastery objectives developed for a given student could be much easier than the set developed for a different student, after taking into account their respective levels of functioning. If, however, each student is assessed on a set of tasks developed to be at the *appropriate level of difficulty*, as the developers of the ALT-MSA intended, mastery percentage comparisons may be appropriate. The goal is for all students to be held to the same standards relative to a set of challenging and appropriate objectives. Therefore, the work or ability required by a student to achieve a 60% mastery percentage (the score needed to be deemed proficient) should be approximately the same for all students regardless of the specific tasks assessed.

Appendix A, Tables 6 and 7 provide mastery percentage frequency distributions in reading and mathematics for the current administration. Average mastery percentage scores are provided in Table 8. In addition, the percentage of students classified in each proficiency level given these mastery percentages can be found in Appendix A, Tables 9-11 and 12-14 for reading and mathematics, respectively. The tables provide counts and percentages for the total group tested, as well as broken out by socioeconomic status (i.e., free/reduced lunch) and ethnicity.

Evidence and Indicators of Positive Practice

After scorers have examined the artifacts submitted with a portfolio for mastery, they review the portfolio as a whole for the presence of certain indicators of best instructional practice for students with significant disabilities.

- Student was involved in the development of the ALT-MSA Portfolio.
- Parent was involved in the development of the ALT-MSA Portfolio.
- The mastery objectives indicate that the student has the opportunity to apply reading and mathematics content standards or access skills to authentic, real-life problems or situations, or other content areas.
- The student's reading and mathematics content standard or access skill objectives reflect age-appropriate materials and tasks.

For each of these indicators a score of zero is assigned if the indicator was not present in the student's portfolio and a score of one is awarded if it was. The percentage of portfolios exhibiting evidence of each of the positive practice indicators is presented in Appendix A, Table 15.

Positive practice data is *not* incorporated into student scores for use in the Maryland Accountability System. The positive practice indicators are intended only to highlight areas for future improvement and to support the link between assessment and instruction. *Please keep in mind that lack of evidence of certain indicators of best practice within a portfolio does not mean that best practice was not followed. It may simply be the case that such indicators are not outwardly apparent or identifiable in the submitted portfolio materials.*

Reports

All districts receive the following standard reports:

Accountability Reports

Home Report

The ALT-MSA home report provides information about an individual's overall performance on the mathematics and reading objectives assessed in the current administration. These reports provide the student's mastery percentage score and corresponding proficiency level for each subject area. The average mastery percentage score for the student's school and district and the state overall is also reported.

Label

A label is produced for each student who participates in the ALT-MSA. The label includes the student's name, gender, ethnicity, LEA, and school name, as well as his/her mathematics and reading proficiency level.

Non-Accountability Reports

Report to Principals

The Principal's report provides a general description of the ALT-MSA program, including the process used to score portfolios and the means by which proficiency level cut-scores were established. This report also provides principals with guidelines for using ALT-MSA scores to support instructional planning and overall program evaluation.

The Principal's report includes a section with student portfolio feedback. This section provides information about a student's performance relative to each mastery objective

assessed. For each mastery objective within a subject area the report indicates whether it was mastered, not mastered, or not scorable. For those mastery objectives deemed not scorable the condition code assigned is provided and defined.

The student portfolio feedback section also presents the assigned score for each of the indicators of positive instructional practice, displaying 0% if the indicator was not present in the student's portfolio and 100% if it was.

School/System/State Summary Report

The format of the school, system, and state summary reports is identical. These reports differ only in the population of students used to calculate the reported results. The summary report provides a general description of the ALT-MSA program, a description of the scoring process, and some guidelines for the use and interpretation of assessment results. In addition to this informative text, two data driven sub-reports are also produced. The first report presents the number and percentage of student portfolios (in the school, system, or state) showing evidence of each of the "Indicators of Important Components of the Instructional Program." The second provides the percentage of submitted artifacts (in the school, system, or state) for mathematics and reading considered mastered, not mastered, and not scorable by grade level.