

Item Development and Review

A primary key to the reliability, validity and utilization of any large-scale test is a rigorous psychometric analysis. Field testing is a necessary step for assessing quality of the items and once the items are placed in the operational test, accurate calibration, equating and scaling should be the most critical process to ensure the reliability and validity of the test scores. Best practices for the MSA Science operational testing were guided by the *Standards for Educational and Psychological Testing* (APA, AERA, & NCME, 1999). Based on these decisions, the following information describes the operational testing of MSA in spring 2008.

Item Development

MSDE and Pearson worked together to define the development targets in support of the 2008 field test. Overall, development was structured to spread the items across the six standards specified within the Maryland Voluntary State Curriculum (VSC) and across the topics, indicators, objectives and assessment limits within each standard. Targets were developed at both grades 5 and 8; item development began once the development targets were finalized. The target number of SR items developed in 2007 for the 2008 administration was 190 items for each grade: 174 SR and 16 BCR items.

During 2006 published technical passages were selected and reviewed by Pearson content staff, MSDE content experts, and three separate Maryland content and bias committees to be approved for item development. An item writer training was held in early January 2007. Current or former non-Maryland Science educators were recruited to write items and lab stimulus on behalf of the program. During the training, writers were introduced to a number of topics by both MSDE and Pearson staff. Topics for training included:

- an introduction to the Maryland VSC
- the concept of assessment limits
- the types of items on the MSA Science test
- elements of universal design in assessment
- how to develop items aligned to standards
- identifying potential bias/sensitivity issues within the materials written, and
- guidelines for writing SR and BCR items.

Following training, writers were given an opportunity to begin drafting items, which were then reviewed by Pearson content staff.

Once Pearson received items from writers, each item underwent an extensive internal review by Pearson content specialists for total item quality, including but not limited to:

- accurate Science content
- appropriate and engaging context
- effectiveness as a measurement of assessment limits within VSC
- age and grade-level appropriate language and vocabulary
- adherence to established MSDE style guidelines

Additionally, Pearson content specialists reviewed all items within each grade for the full range of difficulty and cognitive levels. Items then go through an iterative development process between content specialist and copy editors, universal design specialists, and research librarians. In addition, all art and graphical supports for the items were produced. Finally, all BCR items were reviewed by Pearson Performance Scoring Center staff for scorability. Once Pearson completed the internal development, items were released to MSDE for review via Pearson's Item Tracker system. In May of 2007, Pearson and MSDE content experts met to review and discuss each new item and collaborate on revisions. Once revisions were made and reviewed again through the internal Pearson development team, the items were prepared for another series of content and bias reviews in Maryland.

Review panels of Maryland residents were convened in July 2007. Three different panels were convened to review items for each grade. Content review was conducted at each grade by Maryland educators within the appropriate grade range to further confirm content accuracy and grade-level appropriate vocabulary and language, and to identify and discuss potential improvements to the item stem or distractors. A separate bias/sensitivity panel at each grade was convened to examine the items for any possible socio-economic, geographical, cultural or gender biases. Finally, another committee of educators reviewed item text and graphics with particular focus on possible issues for blind or visually impaired students. Before reviewing materials, MSDE and Pearson provided an overview to the panelists on the purpose of each panel, the VSCs, and the criteria by which they were asked to evaluate the items. Since the evaluation criteria were different, the content panelists and bias/sensitivity panelists were trained separately.

Content panelists were asked to evaluate the materials on the basis of the following criteria:

- alignment to the VSCs
- clarity and grade-appropriateness of text and graphic supports
- accuracy of the underlying Science content

Bias/sensitivity panelists were asked to evaluate the materials as an additional check on whether the materials:

- reflected favoritism towards a gender or ethnic group
- were free of potentially offensive or inappropriate language
- discriminated in any way against individuals who have special needs
- contained any underlying assumptions not shared across ethnic, racial, and gender groups, socioeconomic levels, and geographic areas
- contained language and/or dialect that is not commonly used across the state or has different connotations in different parts of the state
- had graphic supports that were appropriate and accessible for all students

In addition to the panels reviewing the items to be field tested in spring 2008, separate bias and content panels were convened for both grade 5 and grade 8 to read and evaluate the technical passages that were proposed to be used on the spring 2009 embedded field test. On the basis of input from these groups, MSDE and Pearson selected the passages for which items would be developed for the 2009 field test.

Following the panels, MSDE and Pearson met to reconcile the comments from the various groups. Each item and stimulus was reviewed along with the comments from the bias, content

and low-vision panels. From this, a final decision was made by MSDE with respect to all edits and the disposition of the item.