Test Overview and Design

Introduction

The Maryland School Assessment (MSA) tests are measures of students' knowledge relative to the Maryland State Curriculum at grades 5 and 8. The MSA Science test was added to established assessments in Reading and Mathematics to form part of the MSA program. Administered annually in the spring, the MSA program was established to meet the requirements of the No Child Left Behind Act (NCLB) of 2001. In 2006, Pearson was contracted by Maryland State Department of Education (MSDE) to develop, administer, and maintain the MSA Science test. This report provides technical details of work accomplished during the 2011-2012 test administration cycle.

Purpose

The purpose of this MSA Technical Report is to provide objective information regarding technical aspects of the 2012 MSA Science operational test. This volume is intended to be one source of information to Maryland K-12 educational stakeholders (including testing coordinators, educators, parents, and other interested citizens) about the development, implementation, scoring, and technical attributes of the MSA Science tests. Other sources of information regarding the MSA Science test, provided in paper or online format, include the MSA Science administration manual, implementation materials, and training materials.

The information provided here fulfills professional and scientific guidelines for technical reports of large scale educational assessments and is intended for use by qualified users within schools who use and interpret the results of the MSA Science tests. Specifically, information was selected for inclusion in this report based on NCLB requirements and standards from the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 1999).

This manual provides information about the MSA Science test regarding:

1. Content of the tests;
2. Test form design;
3. Identification of ineffective items;
4. Reliability of the tests;
5. Difficulty of the test questions;
6. Equating of test forms;
7. Detection of item bias;
8. Scoring and reporting the results of the tests.

From test development to final reporting, each of these facets of the MSA Science test contributes to the validity of the inferences made about the test results. This technical manual addresses these topics for the 2011-2012 testing year.

Test Overview

In 2002, the Maryland State Department of Education adopted the testing program known as the Maryland School Assessment (MSA). The first two subjects to be established under this new

testing program were Reading and Mathematics. The Science test was added and the first field test administration was conducted in the spring of 2007, followed by the first operational test in 2008. The MSA Science test is currently given to grade 5 and grade 8 students in order to assess achievement in Science. Score reports are provided to parents and include total test scale score results and performance level classifications (described in more detail in following sections).

Purpose and Use

By assessing student achievement against the Science academic standards, the MSA Science test serves two important purposes. First, the MSA Science test provides an accountability tool to measures performance levels of students, schools, and districts against the Science academic standards. Second, it provides parents, teachers, and educators critical information about what students have learned, which, if applied constructively, can foster improvement of instructional programs, classroom education, and school performance.

Test Content, Specifications and Design

The MSA Science test was designed to align to the Maryland State Curriculum (MSC) that specifies curricular indicators and objectives that contributed directly to measuring content standards. According to MSDE's website, the MSC defines what students should know and be able to do and "is the document that aligns the Maryland Content Standards and the Maryland Assessment Program." The MSC is formatted so that content standards delineate broad, measurable statements about what students should know and be able to do. Each standard has multiple indicator statements that provide the next level of specificity, thereby narrowing the focus for teachers further. Finally, objectives provide teachers with very clear information about what specific learning should occur. The MSC is widely disseminated to Maryland educational stakeholders, including teachers, central office staff, students, parents and other stakeholders.

In order to ensure that MSDE is in accordance with the federal law that requires states to align their tests to their content standards, the MSC serves as the guiding document for test development and design. Developing the items for testing was a collaborative effort between MSDE, educators, and Pearson. Teachers, administrators, and content specialists were recruited from all over Maryland for several test development committees. These committees reviewed items developed for MSA Science test.

The basic test specifications were established by MSDE and provided to Pearson to guide the test development and administration. Since the inception of the Science test, there have been six test administrations—a census field test in 2007 and four operational tests (2008 through 2012). All administrations were conducted under the same testing conditions. Accordingly, the field test was designed to match the requirements of the operational administration test blueprint, i.e., a student taking the census field test and the operational test would respond to the same number and type of items. However, because of embedding of field test items on the operational form, there were fewer scored items on the operational form, even with the same number of overall items. Beginning with the 2008 operational test, two base forms (i.e., two forms of scored operational items) were used. Each form had a total of 77 items on the grade 5 form and 75 items on the grade 8 form. Grade 5 tests had 66 operational (yielding a student score) items and 11 field test items for grade 5. The grade 8 test had 64 operational items with 11 field test items. For both grade tests, only operational items contributed to student scores. The two base forms share a set of 20 common items. These common items are discrete (i.e., non-passage based, stand alone) selected response (SR) items.

MSA Science Item Types

The 2012 operational MSA Science included two types of items: selected response (SR) and brief constructed response (BCR). SR items require students to select a correct answer from several alternatives. For the 2012 MSA Science tests, students selected an answer from four options. Each SR item was scored dichotomously (i.e., 0 or 1). BCR items require students to provide a short answer using words, numbers, and/or symbols. All BCR items are scored using a generic rubric and scores range from 0-3 based on concordant scores from two independent raters. In cases where the scores differ by one point, the higher score is used. In cases where the rater scores differ by two or more points, a third expert rater's independent score is used as a resolution.

In addition to these formats, a new item type was administered at the end of the online operational tests. MSDE has been exploring the incorporation of technology enhanced (TE) items for a number of years as a means of potentially measuring more complex skills in line with steps towards Next Generation Assessment. TE items make use of the interactive capacity of computers to allow for enhanced presentation and capture of stimuli and responses. They can range from the simple (i.e. drag-and-drop, hot spot, etc.) to fully interactive multi-step scenario based formats.

Given that MSA Science is currently administered both online and on paper it was important to ensure that inclusion of the TE items was handled in such a way that year-to-year score comparability was preserved. This was addressed by administering a single TE item at the end of the online forms. Additionally, the TE items used were comparable in terms of seat time to complete and complexity to existing SR items.

MSA Science Test Blueprints

There are two MSA Science test blueprints available, one for grade 5 and one for grade 8 and there are six standards assessed across each grade with 66 items in the grade 5 test and 64 items in the grade 8 test, as presented in Tables 1 and 2.

Table 1. Grade 5 MSA Science Standards Assessed

Standard	
1.0	Skills and Processes
2.0	Earth/Space Science
3.0	Life Science
4.0	Chemistry
5.0	Physics
6.0	Environmental
	Total Number of items: 66
	Total number of points:72

Table 2. Grade 8 MSA Science Standards Assessed

Standard	
1.0	Skills and Processes
2.0	Earth/Space Science
3.0	Life Science
4.0	Chemistry
5.0	Physics
6.0	Environmental
	Total Number of items: 64
	Total number of points: 72

MSA Science 2012 Operational Test Construction

The 2012 operational tests were created according to the test blueprints (see Table 1 and 2) and reflective of the Maryland State Curriculum for Science in the form of measureable Indicators and Objectives. As such, each of the two operational forms yielding student scores has the same test composition as that of the 2008 tests in terms of content, total number of items/score points, and item types. Additionally, each operational form was created with five unique sets of embedded field test items (see MSA Science 2012 Field Test Design). As noted in the previous section, the two operational forms were created with a common set of 20 SR items. These items were chosen to reflect a miniature version of the overall operational tests and provide a mechanism for placing all operational items from both forms onto a common scale.

The process of selecting items for the two 2012 MSA Science operational test forms was an iterative process primarily involving Pearson content experts, MSDE, and Pearson psychometricians. Initial test forms were created to meet the respective blueprints, reflect the MSC measureable Indicators and Objectives, and align with statistical characteristics of the 2008 operational tests. Only items deemed eligible after being administered live (field tested) and reviewed by content experts based on statistical indicators (see Data Review of the Field Test Items) were used. Additional content-related characteristics that were part of the creation of the operational test forms had to do with ensuring there was no cuing from one item to the next. That is, items were scrutinized to make sure nothing in any one question or passage would provide information relevant to answering any other item correctly.

Classical item statistics were used in conjunction with item response theory (IRT) statistics to help target the overall test forms. The guiding principles were choosing items with reasonably strong point biserial correlations (ideally >.30) and matching a spread of item difficulties in line with the 2008 forms. Items flagged for any reason based on the data review criteria (also including differential item functioning, as described later) were identified as such, and staff members were discouraged from using them. Item level statistical targets based on overall test, by standard, and by item type were also used for guidance. IRT test characteristic curves (TCCs), test information functions (TIFs), and conditional standard error plots for each test form were also compared to the respective 2008 plots to help ensure the overall IRT measurement properties were captured across the scale (see Test Analysis, Operational Scaling and Scoring).

This process of content and psychometric review and modification of each operational test form proceeded iteratively, where each group would evaluate the most recent proposed forms and provide feedback. Once operational test forms were created that best met all content and statistical targets, the proposed forms were submitted to MSDE for review and/or modification.

MSA Science 2012 Field Test Design

Field test forms were composed of selected response (SR) items and brief constructed response (BCR). Items were either stand-alone (not linked to other items), linked to a lab set stimulus (e.g., technical graph or figure), or linked to a technical passage stimulus. Field test item sets 1-5 were embedded in Form A and 6-10 in Form B. In other words, operational forms 1 through 5 share the same operational items and are differentiated by a unique field test item set within each form. Table 3 presents a graphical representation of this field test design. Items common to both forms are also depicted.

Table 3. 2012 MSA Science Test Form Design

Operational Items	Field test Item Sets									
	1	2	3	4	5	6	7	8	9	10
Form A	X									
		X								
			X							
				X						
Common					X					
Items						X				
Form B							X			
								X		
									X	
										X

MSDE and Pearson worked together to finalize the structure of the 2012 field test forms. At each grade, 10 field test forms were produced. The intent of the test build process was to have each form be parallel in terms of number of SR items, BCR items, and stimulus materials. In addition, the field test forms were designed to be equivalent to the operational base forms plus embedded field test in terms of total numbers of SR and BCR items. All 10 forms per grade had the same number of SR and BCR items. In addition, a goal of item selection was to balance, to the extent possible, coverage of the standards across the 10 field test forms per grade. On a per form basis, initial item selections were performed by Pearson and then shared with MSDE for review and approval. Since Form 1 at each grade was the Braille/large print form, items were selected for Form 1 on the basis of feedback provided by the low-vision panel.

The 2012 forms (and all subsequent operational assessments) were spiraled at the student-level. Spiraling at the student-level supports the assumption that examinee groups responding to each test form are randomly equivalent; an assumption that will further strengthen the link across forms.