Introduction

The 2007 Maryland High School Assessments (MDHSAs) consisted of end-of-course tests in Algebra/Data Analysis, Biology, English, and Government. The MDHSAs are referred to as "end-of-course" tests because students take each test as they complete the appropriate coursework. In addition, results from the English and Algebra administrations are used in the Maryland State Department of Education (MSDE) Adequate Yearly Progress reports, required under the No Child Left Behind Act (NCLB) for the 2007 school year.

The MDHSAs consist of selected-response (SR) items, which require students to choose between four short response options; brief constructed response (BCR) items, which require students to write a short response; and extended constructed response (ECR) items, which require students to write a longer response. In addition, Algebra/Data Analysis tests include items that require a student-produced response (SPR); students must grid in their responses on the answer document. The SR and SPR items are machine-scored; the BCR and ECR items are scored by raters. All items are based on content outlined in Maryland's Core Learning Goals (CLG).

The 2007 MDHSAs were administered in January, May and Summer (July/August). For the January and May administrations, three operational test forms were constructed; one for the primary administration window, and one for each of two make-up administrations. For the Summer administration, two forms were constructed; one for the first week of testing and one for the second week of testing.

Each test form consisted of operational and field test items. Operational items were used to produce student scores. The field test items were scored along with the operational items, but examinees' scores on these items were not included in the computation of a total test score. Instead, the performance on the field test items was analyzed and the calibrated items were placed into the item bank. With the exception of items selected for public release, all operational items are returned to the item bank where they must remain unused for at least two years to minimize item exposure.

The item response models used to calibrate the items in the MDHSA tests were the three-parameter logistic (3PL) model and the two-parameter partial credit (2PPC) model, also known as the generalized partial credit model (GPCM; see Section 5). These models were used to generate both total test scores and subscores. The total test scores were generated using item-pattern (IP) scoring, and the subscores were generated using raw score (RS) to scale score (SS) tables. Total test results in the scale score metric were reported to students. Subscores were not reported to students but were aggregated at the classroom level to provide teachers and administrators with additional information about student performance in each of the subscore categories.

Beginning with the 2004 administration, a pre-equated design was implemented; prior to 2004, scores were based on parameters that were estimated following the administration

using a post-equated[1] design. In the pre-equated design, item parameters are not estimated following an administration; instead existing bank parameters are used to produce student scores. Using this design, scores can be calculated and assigned to students immediately after their answer documents have been processed.

All technical support and analyses were carried out in accordance with both the *ETS Standards for Quality and Fairness* (2002) and the *Standards for Educational and Psychological Testing*, issued jointly by the American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (1999).

The 2007 technical report comprises five sections and two Appendices. Section 1 describes test development, form construction and administration details; Section 2 discusses the validity of the MDHSAs; Section 3 describes the scoring procedures and score types; Section 4 provides statistical summary results for each of the test forms administered in 2007; and Section 5 describes the analysis of the field test data including classical item analysis, differential item functioning, and item response theory calibration and scaling. Appendix A provides classical item statistics for each operational test item by form administered in 2007. Appendix B provides classical item statistics for each field-test item administered in 2007.

---

[1] In the post-equated design, anchor items representative of the content and difficulty of the test forms were used to equate the test forms using a Stocking and Lord procedure (CTB/McGraw-Hill, December, 2003).