

Introduction

The 2004 Maryland High School Assessments (MHSA) consisted of end-of-course tests in Algebra/Data Analysis, Biology, English I, Geometry, and Government. The HSA is referred to as “end-of-course” tests, because students took each test as they completed the appropriate coursework. In addition, results from the Geometry administration were used as a High School mathematics component in the Maryland State Department of Education (MSDE) adequate yearly progress reports as required under the No Child Left Behind (NCLB) act. HSA contained selected-response (SR) items, which required students to choose between three/four short response options and are machine scored; brief constructed response (BCR) items required students to write a short response and are scored by raters; extended constructed response (ECR) items required students to write a longer response and are also scored by raters. In addition, Algebra/Data Analysis and Geometry included items based on student-produced response (SPR), which required students to grid in correct responses to the answer document. All items were based on content outlined in Maryland’s Core Learning Goals.

HSAs were administered in January, May and July. In general, for January and May 2004 administrations, three operational test forms were constructed: one for the main administration window, and one for each of two make-up administrations. In order to conserve the item pool, two May make-up forms were used for July main (May make-up form 1) and July make-up form (May make-up form 2). Each test form consisted of two types of items: operational and field test. Operational items were common across each of the operational forms and were used to produce student scores; field test items were not scored operationally, but were analyzed and placed into the item bank for future test form construction. In addition, with the exception of items selected for public release, all operational items were also returned to the item bank where they are to remain unused for at least two years to minimize item exposure.

The underlying item response models used for HSA 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). For each content area, both a total test score and subscores were reported to students. The total test scores were reported to individual students and were based on item-pattern (IP) scoring (mean 400, standard deviation 40). Subscores were also reported based on associated item parameters, though these scores were obtained using number-correct (NC) to scale-score (SS) tables. A study was conducted to investigate the nature and extent of differences in subscores based IP scoring versus NC scoring (see Chapter 3). While subscores were not reported at individual student level, the subscores were aggregated at classroom level to provide teachers and administrators with additional information about student performance in each of the reporting categories. A special study was also conducted that involved reviewing and replicating English 2003 results using ETS programs. Results indicated that the ETS programs successfully replicated the English 2003 results reported by CTB (see Chapter 3).

Beginning with the 2004 administration, a pre-equated design was implemented while scores from previous administrations were based on parameters that were estimated following the administration (post-equated¹). In the pre-equated design, item parameters were not updated following an administration; instead existing bank parameters were used to produce student scores. Using this design, scores can be calculated and assigned to students immediately after the answer documents have been scored.

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

This report is divided into 5 sections: Section 1 describes test development, form construction and administration details; Section 2 discusses the validity and reliability of the HSAs; Section 3 describes the scoring procedures and score types; Section 4 provides statistical summary results for each of the test forms administered in 2004; and Section 5 describes the analyses conducted using the field test data including classical item analyses, differential item functioning, and item response theory calibrations and equating.

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).