Section 4. Scoring Procedures

Scale Scores

The MD HSA reporting scale ranges from 240 to 650. For Algebra, Biology, and Government, the scale was established in 2003 and defined so that the scale scores had a mean of 400 and a standard deviation 40. In 2005 a scale for English was established that had the same range, mean, and standard deviation.

These scores represent ability estimates obtained using Item Response Theory (IRT, Yen & Fitzpatrick, 2006). (See IRT Calibration and Scaling in Section 7 for details about the three-parameter logistic model (3PL) used for the MD HSAs.) Students' total test scores and subscores are scale scores derived using the 3PL model and item-pattern (IP) scoring procedures. When the 3PL model is used, the likelihood equation can have multiple maxima. Therefore, a numerical method was developed to find the scale score at the global maximum in the likelihood function (Yen, 1984).

Conditional Standard Errors of Measurement

Corresponding conditional standard errors of measurement (CSEM) were produced and are equal to the inverse of the square root of the test information function.

$$CSEM(\hat{\theta}) = \frac{1}{\sqrt{I(\hat{\theta})}}$$

where CSEM($\hat{\theta}$) refers to the conditional standard error of measurement and I($\hat{\theta}$) refers to the test information function for $\hat{\theta}$. The test information function is the sum of corresponding information functions of the test items when optimal item weights are used. Item information functions depend on the item difficulty, discrimination, and conditional item score variance.

Lowest and Highest Obtainable Test Scores

The maximum likelihood procedure under the 3PL model cannot produce reasonable scale score estimates for students with perfect scores or scores below the level expected by guessing. While maximum likelihood estimates are usually available for students with extreme scores other than zero or perfect, occasionally these estimates have very large CSEMs, and differences between these extreme values have little meaning. Therefore, scores were established for these students based on a rational procedure (refer to Appendix 3.C of the 2004 Technical Report). These values were called the lowest obtainable scale score (LOSS) and the highest obtainable scale score (HOSS).

Cut Scores

MSDE established the cut scores associated with each of the performance levels in the content areas other than English in 2003.[6] The English cut scores were established during the standard-setting meeting held in October 2005. One cut score was established for Biology and one was established for Government. Because Algebra and English results are used as the high school mathematics and English/language arts components of the Maryland accountability plan under NCLB, two cut scores were established for these content areas. To comply with NCLB requirements for secondary science, an Advanced cut score for Biology was established in 2008. These values are given in Table 4.1.

Table 4.1 MD HSA Cut Scores by Content Area

	Cut-Score	
Content Area	Proficient	Advanced
Algebra	412	450
Biology	400	452
English	396	429
Government	394	

Beginning with the class of 2009, students must obtain either a passing score on each of the four MD HSAs or an overall combined score of 1602 as part of the requirements for a high school diploma. Passing status is achieved when a student's score meets or exceeds the Proficient cut score, as listed in Table 4.1. Students graduating prior to 2009 were not required to pass the MD HSAs but were required to take the tests.

Year-to-Year Scale Maintenance

The MD HSAs for Algebra, Biology, and Government have been pre-equated since 2004; English has been pre-equated since 2005. In the pre-equated design, a pool of IRT-calibrated items expressed on the reporting scale exists for test form construction. The item parameter estimates for new forms are obtained from the bank and are used to build test forms that are parallel across administrations. Student scores are produced with the bank-obtained item parameter estimates, thereby linking scores from one administration to the other.

To increase the item pool, the MD HSA embeds field test items in the operational test forms. The field test data for the January and May administration are calibrated with the operational items at that time. The calibrations are linked to the reporting scale using all operational items as anchors and the Stocking and Lord procedure (Stocking & Lord, 1983). Having all operational items serve as linking items ensures that the linking set is large enough to provide stable and reliable

[6] Technical documentation on the standard-setting method used to establish the MD HSA cut scores is available on the Maryland State Department of Education website at http://www.marylandpublicschools.org/msde/divisions/planningresultstest/maryland+standard+setting+technical+reports.htm.

results. Item bank parameter estimates are established at the time of the field test and are not updated following each administration.

To ensure that items behave the same way across administrations, construction of new forms follows guidelines defined by Kolen and Brennan (1995). These guidelines are:

1. Items should appear in the same contexts and positions as when the item parameter estimates were established. Operational items are placed as close as possible to the same position they were in when parameters were estimated and within the same third of the total test form.
2. Operational items should appear in similar positions on the test. It may be problematic if an item is positioned in very different locations on the two forms, such as at the beginning of the test on one form and at the end of the test on another form. Operational items that appear in more than one form occupy consistent positions across forms; MSDE must approve any deviations.
3. The text is exactly the same in the old and new forms. Minor editorial changes and rearranging answer choices are discouraged; otherwise the items may function differently. All requests for minor editorial changes must undergo psychometric review to evaluate the implications for the response process.