Section 3. Scoring Procedures and Score Types

Scale Scores

Scale scores based on maximum likelihood estimates (MLE) were reported for the total test score. All scores were reported on the operational reporting scale established in 2003. While the total test score was based on item-pattern (IP) scoring, the subscores were based on number-correct (NC) to scale score scoring tables.

With IP scoring, because the likelihood equation can have multiple maxima with the 3PL model, a numerical method was developed that found the scale score at the global maximum in the likelihood function. NC to scale score scoring tables were obtained by inversing the test characteristic curves (TCC) of items contributing to the associated subscores. The procedure produced what Yen (1984) called 'number correct trait estimates,' which is referred to as 'NC scale scores' in this report.

Conditional Standard Errors of Measurement

Corresponding conditional standard errors of measurement (SEM) were also produced for both types of scoring and were equal to the inverse of the square root of the test information function.

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

where,

$SEM(\hat{\theta})$=standard error of measurement

$I(\theta)$= test information function.

The test information function is the sum of corresponding information functions of the test items when optimal item weights are used, as in the MHSAs. Item information functions depend on the item difficulty, discrimination and conditional item score variance. Thus, while polytomous items often have lower discriminations than selected response items (Fitzpatrick et al., 1996), they may convey more information than selected response items, because they have more score points.

Lowest and Highest Obtainable Test Scores

Both the maximum likelihood procedure and NC scoring cannot produce scale score estimates for students with perfect scores or scores below the level expected by guessing. Also, while maximum likelihood estimates were available for students with extreme scores other than zero or perfect, occasionally these estimates have very large conditional SEMs, 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). The same LOSS and HOSS values were used for either number-correct (NC) or item-pattern scoring. Table 3.1 lists the LOSS and HOSS values for each content area established at the beginning of the MHSA program. MSDE decided that the LOSS and HOSS values for the Summer of 2005 and subsequent administrations would be 240 and 650, respectively, for all content areas.

Table 3.1 LOSS and HOSS Values

Content	LOSS	HOSS
Algebra	240	625
Biology	260	650
English I	240	650
Geometry	275	575
Government	260	650

Cut-Scores

The cut-scores associated with each of the performance levels in the non-English content areas were established by MSDE in 2003 (refer to Table 3.2). The English cut-scores were established during the standard setting study held in October of 2005. One cutscore was established for all of the content areas except for Geometry and English. Because Geometry and English results are used as the High School Mathematics and English Language Arts components of the MD accountability plan under NCLB, two cut-scores were established.

Table 3.2 MHSA 2005 Cut-Scores

	Cutscore	
Content Area	Proficient	Advanced
Algebra	412	
Biology	400	
Geometry	411	447
Government	394	
English	396	429