Section 2. Test Construction and Administration

Test Development

Planning

Planning for the test development process began with the creation of item development plans for each content area. ETS content leaders collaborated with their content counterparts at MSDE to create these plans. The item bank was reviewed to determine how well the available item pool matched the test form requirements set forth in the test form blueprint as defined by the Core Learning Goals (CLGs) and the 2011–2012 form construction templates provided by MSDE. Areas that contained low item counts were given priority when determining which indicators were to be addressed by the item writers. After these areas with critical need were defined and addressed, any remaining items to be developed (as determined by the requirements set forth in the RFP) were distributed among the indicators in a fashion that would best ensure sufficient numbers of items for use in the construction of forms for future administrations.

Item Types

As noted in Section 1, two item types were included on the 2011-2012 MD HSA tests. These item types include the following:

- Selected response (SR)—questions in multiple-choice format with four answer options;
- Student-produced response (SPR)—an item type used in Algebra only, for which the student works the problem and records the solution in an answer grid.

Table 2.1 shows how these item types and associated points were distributed by content area. Each SR and SPR item is worth one point.

Table 2.1 Number of Operational Items by Item Type for Each MD HSA Content Area

Content Area	Operational Items by Item Type		
	SR	SPR	Total
Algebra	43	10	53
Biology	76	-	76
English	60	-	60

Test Specifications and Design

MSDE predetermined the basic test design and provided it to ETS in the form of the content-specific "Test Specifications—Test Form Matrix" document. This basic test design document provides information based on specified expectations and the distribution of the number of items by item type for each reporting category. How the specific items were placed throughout the forms was left to the collaborative efforts of ETS and MSDE content specialists. Construction of the forms was based on test blueprints approved by MSDE. Blueprints for each content area are presented in Tables 2.2 to 2.4.

Table 2.2 MD HSA Algebra Blueprint

	Number of Items		Total Points per Category
	SR (1 pt)	SPR (1 pt)	
Expectation 1.1 Analyzing Pattern and Functions	11	2	13
Expectation 1.2 Modeling Real-World Situations	13	4	17
Expectation 3.1 Collecting, Organizing and Analyzing Data	8	4	12
Expectation 3.2 Using Data to Make Predictions	11	0	11
Total	43	10	53

Note: Information on the referenced indicators can be found in the Maryland Core Learning Goals for Algebra, available on the Maryland School Improvement website at http://www.mdk12.org/assessments/standards/9-12.html

Table 2.3 MD HSA Biology Blueprint

	Number of Items	Total Points per Category
Goal 1 Skills and Processes of Biology	16	16
Expectation 3.1 Structure and Function of Biological Molecules	12	12
Expectation 3.2 Structure and Function of Cells and Organisms	13	13
Expectation 3.3 Inheritance of Traits	13	13
Expectation 3.4 Mechanism of Evolutionary Change	9	9
Expectation 3.5 Interdependence of Organisms in the Biosphere	13	13
Total	76	76

Note: Information on the referenced indicators can be found in the Maryland Core Learning Goals for Biology, available on the Maryland School Improvement website at http://www.mdk12.org/assessments/standards/9-12.html

Table 2.4 MD HSA English Blueprint

	Number of Items	Total Points per Category
1: Reading and Literature: Comprehension and Interpretation (RC) Includes indicators 1.1.1, 1.1.2, 1.1.3, 1.2.1, 1.3.3, 3.2.2	16	16
2: Reading and Literature: Making Connections and Evaluation (RE) Includes indicators 1.1.4, 1.2.2, 1.2.3, 1.2.4, 1.2.5, 1.3.5, 4.2.1	14	14
3: Writing: Composing (WC) Includes indicators 2.2.1, 2.2.2, 2.2.3, 2.2.5, 2.3.1, 2.3.3, 4.3.1	16	16
4: Writing: Language Usage and Conventions (WL) Includes indicators 3.1.3, 3.1.4, 3.1.6, 3.1.8, 3.3.1, 3.3.2	14	14
Total	60	60

Note: Information on the referenced indicators can be found in the Maryland Core Learning Goals for English, available on the Maryland School Improvement website at http://www.mdk12.org/assessments/standards/9-12.html

Item Writing

Item writers were employed to develop high-quality test items that were aligned with the Core Learning Goals. Nearly all item writers were Maryland educators. Only a small portion of the total number of items written was developed by ETS content specialists. Item writers were selected on the basis of their depth of content knowledge and familiarity with the MD HSA program. Many were experienced MD HSA item writers.

Item writers were trained on general item writing techniques as well as writing guidelines that are specific to the MD HSA program. Approximately one month after the initial item writer training, a follow-up training session was provided. The session was designed to evaluate how well the item writers' writing skills had developed to that point, to facilitate peer review of items, and to provide constructive feedback to guide the rest of their writing assignment.

Upon completion of their writing assignment, item writers submitted their items to ETS. Items that were accepted proceeded to the item review and revision process.

Item Review and Revision

All items underwent a series of editorial reviews in accordance with the following procedures:

- Items were edited according to standard rules developed in conjunction with MSDE.
- Items were reviewed for accuracy, organization, comprehension, style, usage, consistency, and fairness/sensitivity.
- Item content was reviewed to establish whether the item measured the intended Goal-Expectation-Indicator-Assessment Limit, with the Goal being the broadest category and Assessment Limit being the narrowest parameter of content being assessed.
- Copyright and/or trademark permissions were verified for any materials requiring permissions, for both field test and operational material.
- Internal reviews were conducted and historical records were established for all version changes.

After ETS performed the required internal reviews, items were submitted to MSDE for review. If the MSDE content specialist requested an original version of the item as submitted by the item writer, a copy was provided. Any associated stimulus material, graphic, and/or art was provided as well as information regarding the Goal-Expectation-Indicator-Assessment Limit that each question addressed.

MSDE content specialists performed a review of the items and provided feedback to ETS content specialists. Edits resulting from this review were incorporated into the items. MSDE and ETS content specialists then met to conduct a side-by-side review of the items. Any further edits to the items based on this review were made. Finally, the items were prepared for review by the Content and Bias/Sensitivity Review Committees.

The Content and Bias/Sensitivity Review Committees are diverse groups of Maryland educators who reviewed each item to ensure that its content (a) accurately reflected what was taught in Maryland schools; (b) correctly matched the intended CLG indicator; and (c) did not unfairly favor or disadvantage an individual or group.

Upon completion of this final round of reviews, MSDE and ETS content specialists conducted another side-by-side meeting to evaluate the reviews and to reconcile the results of the various groups. ETS then made the requested edits to the items and/or revisions to the accompanying graphics. The items that survived this process were eligible for placement in the field test sections of the test forms.

Testing Accommodations

A number of alternate test formats are available to MD HSA examinees, including large-print, Braille, online audio, and Kurzweil versions of the MD HSA developed for each content area. All four alternate test formats are available at each administration. Data from these alternate formats are included in the psychometric analyses.

Test Specifications

All 2012 test forms were constructed using items from the Maryland item bank. The pool of items available for use in the construction of the 2012 forms included all items that had been administered, calibrated, and linked to the operational scale. For Algebra and Biology, the MD HSA operational scale was defined in 2003 and included items administered in 2002 and 2003. For English, the scale was redefined in 2005 when the English test was updated to become an end-of-course assessment for English II. Items flagged for poor fit were excluded from the item pool. Items flagged for substantial differential item functioning (DIF) against one of the focal groups are marked as such in the item bank and they are not used unless required to fulfill content specifications, and then, only after review and approval by MSDE. (See Section 7 for a more detailed account of these analyses and flagging criteria.)

Each test form was constructed to meet specific test blueprint specifications. Tables 2.2 through 2.4 starting on page 13 indicate the distribution of items within each reporting category by item type and the number of score points associated with each item type.

Item Selection and Form Design

To conserve the item pool, when multiple forms were included in an administration, each test form consisted of a common set of operational items shared across forms within an administration, as well as a unique set of items. Within these administrations (i.e., January, May, and Summer), approximately 60 percent of the operational items in each form were common across the test sections. The remaining items in the forms consisted of combinations of items that varied across forms. The guidelines used to construct the forms are provided in Tables 2.5 to 2.9. The exact composition of the forms varied slightly based on available items in the pool.

Table 2.5 Form Construction Specifications for the MD HSA October 2011 Administration

Primary Week Form R
Operational and Field Test items— Reuse of intact form from a prior administration

Table 2.6 Form Construction Specifications for the MD HSA January 2012 Administration

Primary Week Form A	Primary Week Form B	Makeup 1 Form C
Common set—60% Unique items—40% Field Test Section— Unique items	Same as Form A Same as Form A Field Test Section—Unique items	Common set—60% Unique items—40% Field Test Section—Reuse of field test set from Form A or B, or a combination of both

Table 2.7 Form Construction Specifications for the MD HSA April 2012 Administration

Form S
Operational and Field Test items— Reuse of intact form from a prior administration

Table 2.8 Form Construction Specifications for the MD HSA May 2012 Administration

Primary Week Forms D–N	Makeup 1 Form X	Makeup 2 Form Y
Common Set—60%	Common Set—60%	Common Set—60%
Unique Items—40%	Half of items from primary week's 40% unique items—20% Unique items—20%	Other half of items from primary week's 40% unique items—20% Unique items—20%
Field Test Section—Unique sets of items for Forms D–N	Field Test Section—Reuse of one or a combination of the field test sets used in forms D–N, with a preference for form D; however, the actual selection of field test items was determined by the constraints imposed by the operational items	Field Test Section—Reuse of one or a combination of the field test sets used in forms D-N, with a preference of using the same set used for form X; however, the actual selection of field test items was determined by the constraints imposed by the operational items

Table 2.9 Form Construction Specifications for the MD HSA Summer 2012 Administration

Primary Week 1 Form P	Primary Week 2 Form Q
Common Set—60% Unique items—40%	Common Set—60% Unique items—40%
Field Test Section—Reuse of prior administration field test set	Field Test Section— Reuse of prior administration field test set

In addition to the operational items, embedded field test items were included with each version of the test form, resulting in multiple versions of a test form containing different sets of field test items. The percentage of field test items per form varied by content area and administration, as shown in Table 2.10.

Table 2.10 Number of Operational (OP) and Field Test (FT) Items by Administration and MD HSA Content Area

Content Area	OP Items	FT Items	% FT Items
Algebra	53	16	23%
Biology	76	23	23%
English	60	38	39%

Items being field tested were primarily newly written items, with a small number of previously administered items that had been revised due to content concerns or problematic item statistics. Items with problematic statistics were ones that were judged to be acceptable from a content perspective but had one or more of the following statistical characteristics: *p*-values less than 0.10; item-total correlations of less than 0.15; very high omit rates (5% or more); or SR items with a positive point-biserial correlation for one or more distractors. For administrations in which there was more than one primary form available at the same time (January and May), the forms were spiraled at the student level. Spiraling at the student level means that multiple forms of the test were packaged in order (e.g., D, E, F, etc.) and distributed to students according to this order. Spiraling at the student level helps ensure that all forms are randomly distributed throughout the state.

Forms were constructed using the test construction software associated with the customer item bank. The goal was to match the test characteristic curves (TCCs) and the conditional standard error of measurement (CSEM) curves with the "target" form defined as the base form used to set the operational scale. For Algebra and Biology, the base forms were originally developed in 2003; for English the base form was originally developed in 2005. These base forms contained constructed response (CR) items. However, in 2009, CR items were discontinued on the MD HSAs. Because of this change in test design, the target TCCs for the HSAs were revised so that they were no longer influenced by the characteristics of CR items. (Refer to ETS memorandum: *Considerations for Setting New Target Test Characteristic Curves for the Maryland High School Assessments (HSAs)*, October 5, 2009 for details on how new target TCCs were created.)

The following general steps were completed during the test construction process:

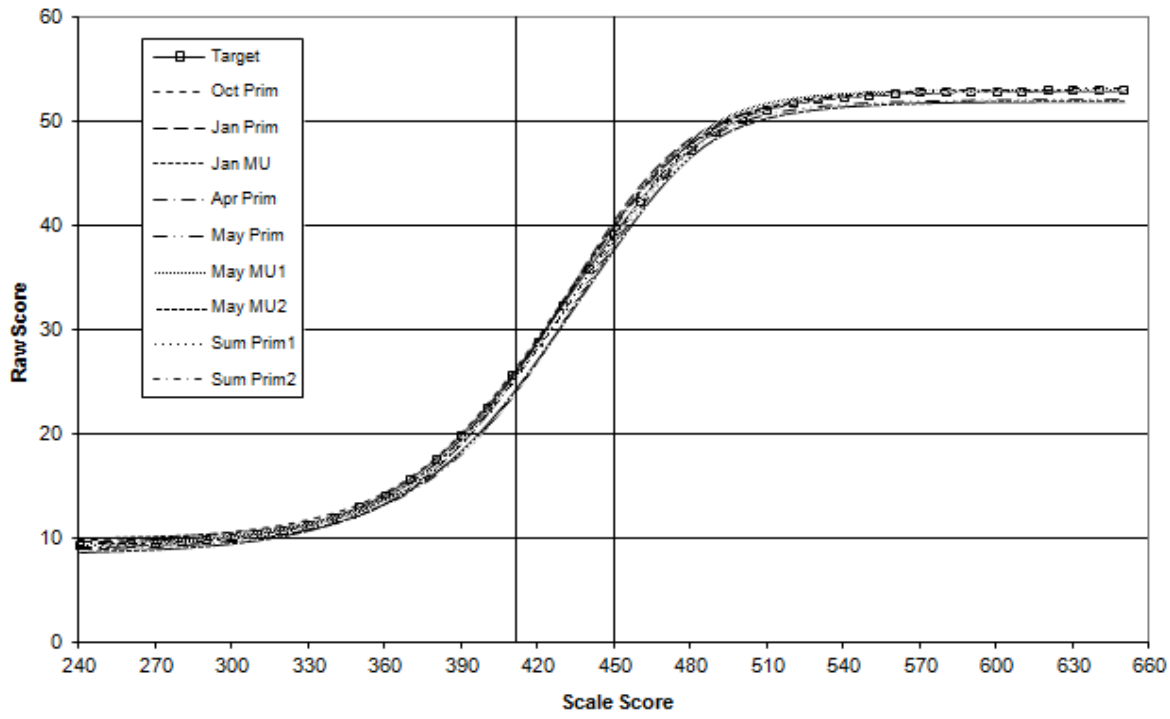
1. For each administration, all forms were constructed simultaneously in order to provide the best opportunity to construct parallel forms.
2. Items that matched the test blueprint were selected to match the target TCCs and CSEMs.
3. Test developers were careful to ensure that the item selections met all content specifications, including matching items to the test blueprint, distribution of keys, and avoidance of clueing[3] or clanging.[4]

[3] Clueing refers to information within a passage, stimulus, item, graphic, or other test component that allows respondents to select/construct the correct answer to one or more items in an assessment without the knowledge and/or skill targeted by the item.

4. After the operational items were selected for the test forms, the field test sets were constructed. Field test sets consisted of SR items in all content areas as well as SPR items for Algebra only. While the field test sets were not constructed to meet any psychometric criteria, they were constructed to meet content criteria. For Algebra and Biology, the field test sets were estimated to be able to be completed by students in approximately 30-35 minutes. For English, the field test sets were estimated to be able to be completed by students in approximately 60 minutes due to the additional time required to read the passages and stimuli. The field test items were embedded in the test according to a variety of content and template criteria, including, but not limited to, coverage of the reporting categories and assessment limits, cognitive balance, key balance/distribution, and clueing/clanging within the field test set and among the surrounding operational items.

Figures 2.1 to 2.6 show the plots of the TCCs and CSEMs for the target form and forms developed for each content area. It is important to note that the TCCs and CSEMs shown in the plots are based on preequated item parameters and therefore are curves calculated prior to administration of the tests. In general, the TCCs and CSEMs were similar to the target curves. The TCC plots indicate that all forms for each content area were within or very close to the acceptable range of the target curve for the full range of scale score values. Note that in the April administration, intact forms from previous administrations were used. Those forms were built to the previous target TCCs, so in some cases deviate more from the current target TCC than is typically seen. Where forms varied in difficulty, differences between forms were typically less than 5 percent of the total raw score across the score range, especially in the range of the cut scores. Where forms had differences slightly greater than 5 percent, these larger differences were typically seen at the very low end of the scale score range and at the high end of the scale. As expected, the CSEM plots indicate that the CSEMs for each content area were lowest in the middle range of scale scores, where the majority of student scores are located. (Please refer to Figures 6.1 to 6.3 for histograms of student performance.)

[4] Clanging occurs when an identical or resembling word(s) appears in both the item stem and one or more item distractors. Also, if two or more items that are near each other share common key words, even if the item content does not clue, the items are said to clang because the interpretation of the word in one item can affect the interpretation of another item.



Note: Maximum possible raw score is 53.

Figure 2.1 Test Characteristic Curves for the MD HSA 2012 Algebra Forms

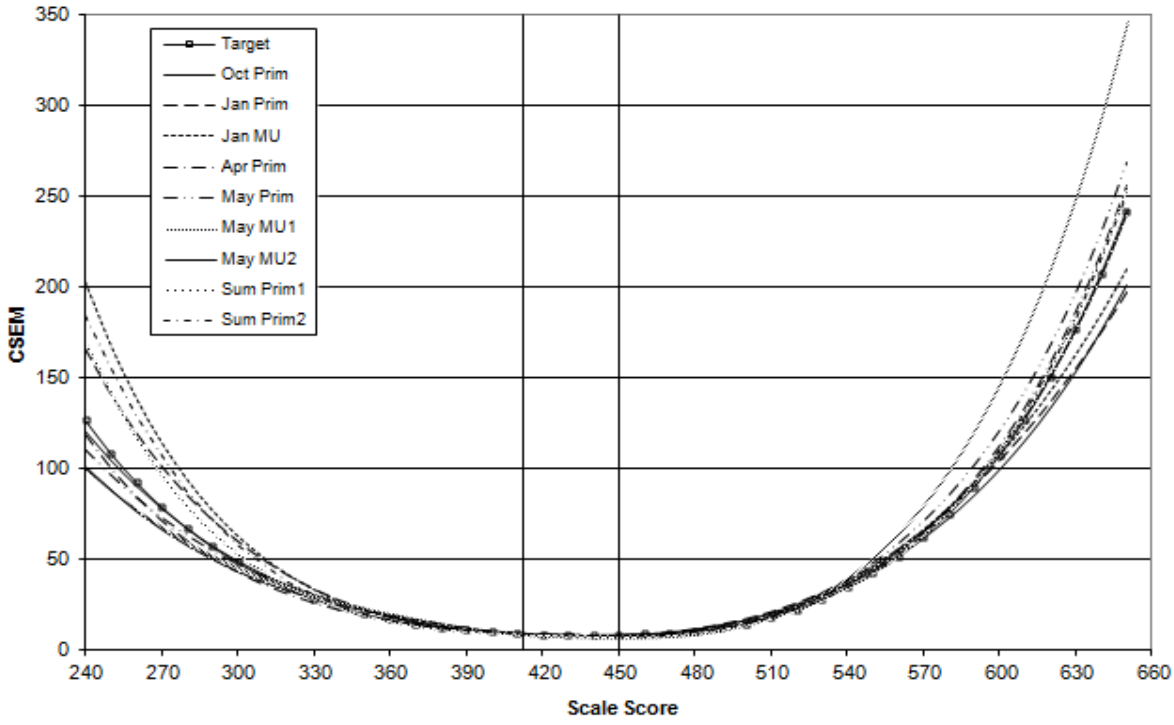
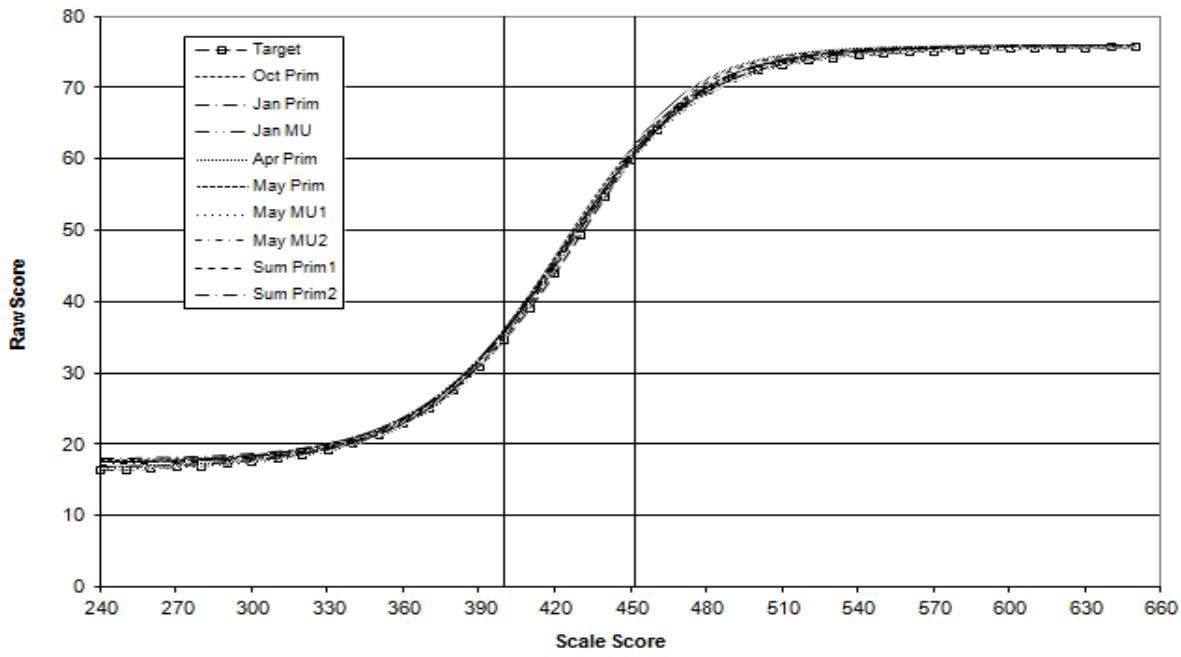


Figure 2.2 Conditional Standard Error of Measurement for the MD HSA 2012 Algebra Forms



Note: Maximum possible raw score is 76.

Figure 2.3 Test Characteristic Curves for the MD HSA 2012 Biology Forms

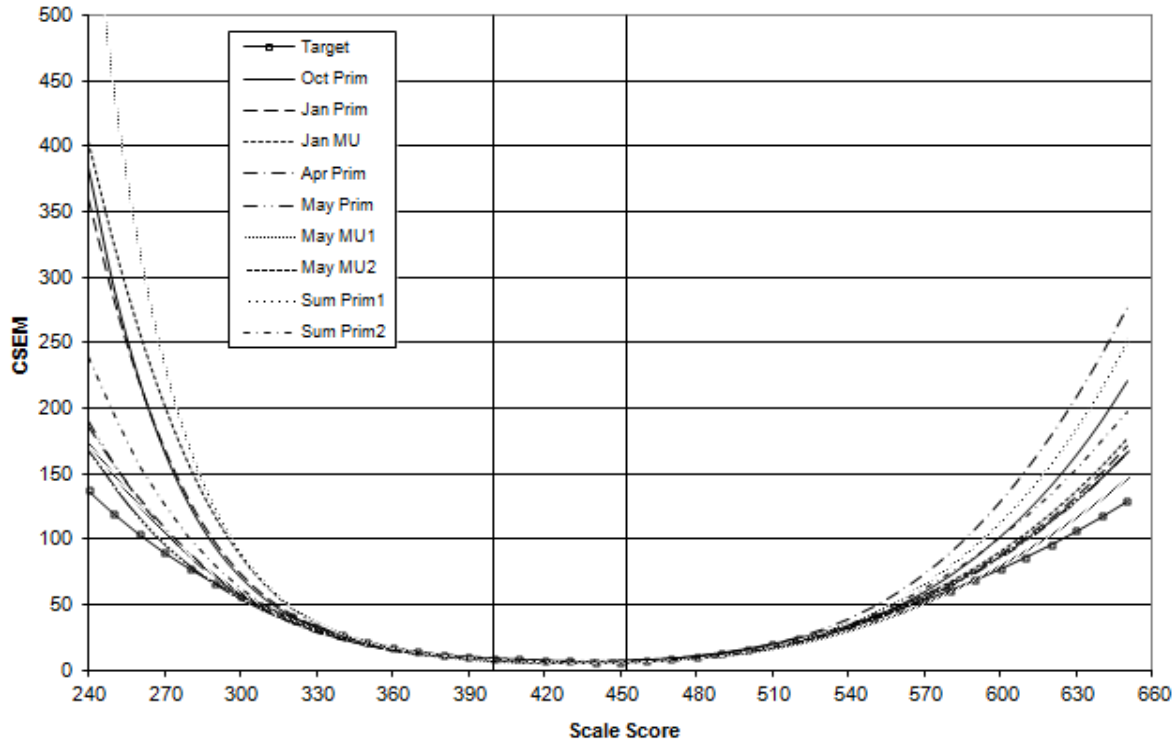
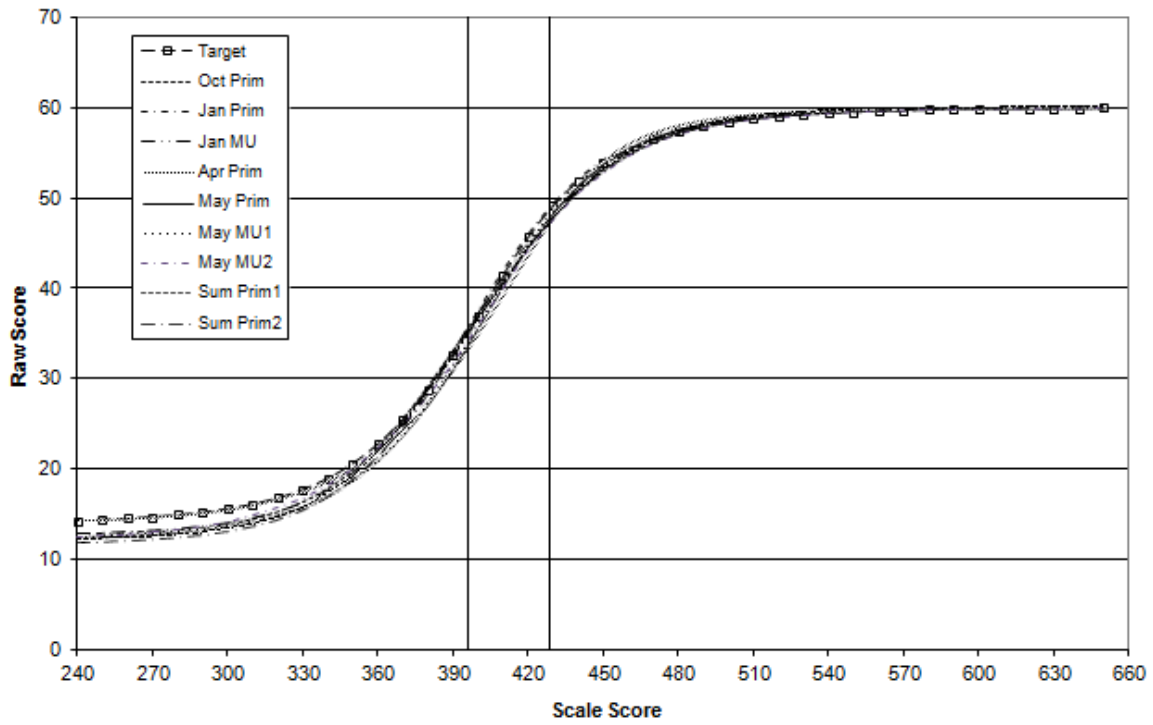


Figure 2.4 Conditional Standard Error Measurement for the MD HSA 2012 Biology Form



Note: Maximum possible raw score is 60.

Figure 2.5 Test Characteristic Curves for the MD HSA 2012 English Forms

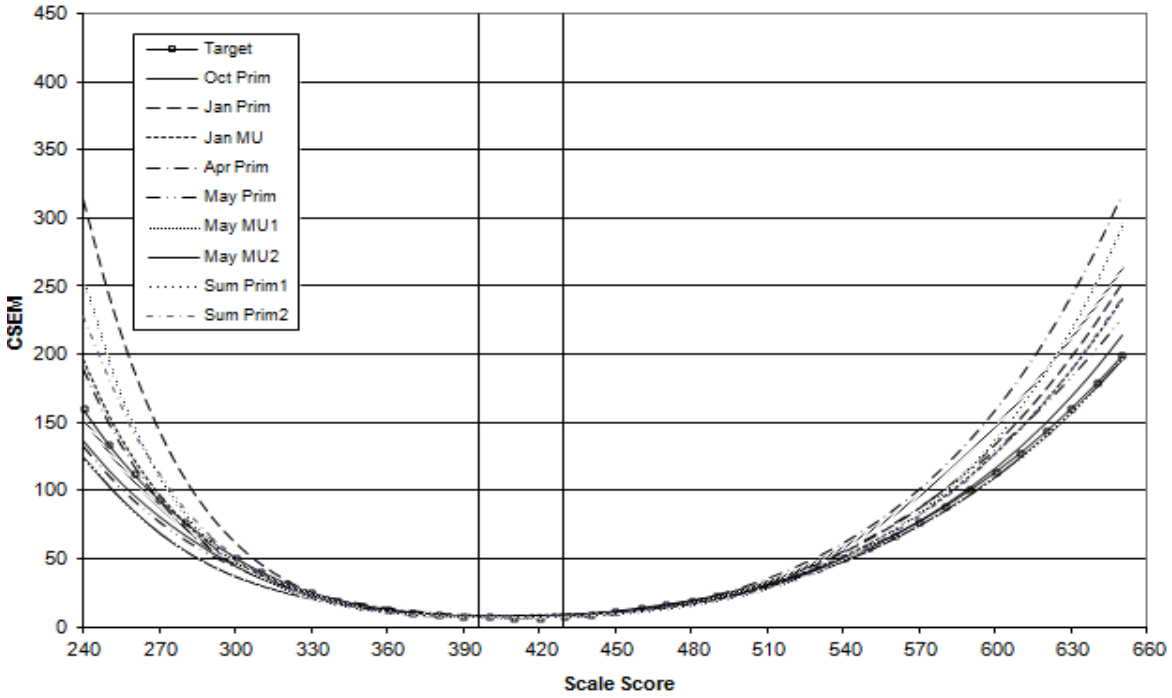


Figure 2.6 Conditional Standard Error of Measurement for the MD HSA 2012 English Forms

Test Administration

For all MD HSA tests administered in 2011-2012, both paper-and-pencil and online versions were available.

For all administrations, paper-and-pencil primary forms were given during the first week of testing. For the January and May administrations, Makeup Form 1 was offered during the second week. For the May administration only, Makeup Form 2 was administered in the third week of testing.

For the online versions in all administrations, the primary and makeup forms were spiraled equally throughout the testing window. In October, only one form was administered. In January, the two primary forms and one makeup form were spiraled over the two-week testing window. In April, only one form was administered. In May, the ten primary and two makeup forms were spiraled over the three-week testing window. In Summer, the two primary forms were spiraled over the two-week testing window. All forms administered without extended time accommodations had timing limits indicated in Table 2.11. The percent of students taking HSAs online by content area and administration is reported in Table 2.12.

Table 2.11 Test Timing Schedule in Minutes by MD HSA Content Area

Content Area	Session One	Break	Session Two	Break	Session Three
Algebra	50	5	50	5	50
Biology	45	5	45	5	45
English	50	5	55	5	50

Table 2.12 Percent of Online MD HSAs by Content Area and Administration

Content Area	October	January	April	May	Summer
Algebra	42%	52%	39%	35%	75%
Biology	45%	54%	47%	46%	66%
English	47%	48%	44%	41%	79%