2020 Technical Manual for the Government and Maryland Integrated Science Assessments (High School Level)

cognia™

Foreword

The technical information included in this report is intended for use by those who evaluate tests, interpret scores, or use test results in making educational decisions. It is assumed that the reader has some technical knowledge of test construction and measurement procedures, as stated in *Standards for Educational and Psychological Testing* (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014).

Table of Contents

List of Tables

List of Figures

Section 1. Introduction

The Maryland High School Assessments (HSAs) are tests that cover core academic areas in Science and Government. The HSAs consist of an end-of-course exam in Government and a cumulative exam in Science, the High School Maryland Integrated Science Assessment (HS MISA). The HSAs are intended to meet the testing requirements for Maryland high school graduation. The HS MISA meets the high school testing requirements for the federal Every Student Succeeds Act of 2015 (ESSA). The HSA Government exam meets the high school testing requirements from Maryland Code Educational Article §7-203 Education Accountability Program 2017. This report provides information about the January 2020 administrations for the HSA Government and HS MISA. Due to the COVID-19 global pandemic the May 2020 administration of HS MISA and May and Summer 2020 administrations of HSA Government were not conducted.

The Government test administrations began in 2002 and continued until 2011. From summer 2011 to October 2012, the Government test was excluded from the Maryland HSAs. Starting in January 2013, the Government test was reintroduced into the Maryland HSAs. HSA Government is referred to as an "end-of-course" test because students take it as they complete the appropriate coursework, while HS MISA is an integrated assessment taken at the end of a locally decided sequence of courses. Starting in 2018, the HS MISA, a high-school level science assessment that is aligned to the Next Generation Science Standards (NGSS), replaced the existing end-of-course assessment in Biology.

Starting in 2016, the end-of-course tests in Algebra and English were replaced by Partnership for Assessment of Readiness for College and Careers (PARCC) assessments. Students who were enrolled in HSA-aligned courses (Government and Biology) during the 2016-2017 school year were required to pass the HSA, achieve an approved combined score[1], or satisfy the graduation requirement via the Bridge Plan[2]. Students entering ninth grade in school year 2013–14 and beyond must pass HSA Government, achieve an approved combined score, or satisfy the graduation requirement via the Bridge Plan. The combined score options varied, depending on whether students have a score from the previous HSA English or HSA Algebra assessments. Students taking the HS MISA in 2020 were not required to pass the HS MISA but were required to participate in the HS MISA to meet the graduation requirement[3] of a Science assessment.

Since May 2009, the Maryland HSAs have been administered online as well as in the paper-and-pencil format. Studies of the comparability of online and paper forms of the Maryland HSAs were conducted in 2009 and 2010. The 2009 report is provided in the 2009 HSA Technical Report in Appendix 1C. The 2010 results were provided to the Maryland State Department of Education (MSDE) (Educational Testing Service, October 29, 2010). Further mode comparability studies have not been conducted.

---

[1] More information on the Combined Score Option is available on the Maryland State Department of Education Website at:
http://www.marylandpublicschools.org/programs/Documents/Testing/GraduationsRequirements2018.pdf

[2] The Bridge Plan provides a process that helps ensure all students have a fair opportunity to demonstrate their knowledge and skills if traditional testing instruments are not effective measures for them. See more details at:
http://www.marylandpublicschools.org/programs/Documents/Testing/GraduationsRequirements2018.pdf

[3] More information on the testing requirement for graduation is available on the Maryland State Department of Education Website at:
http://www.marylandpublicschools.org/programs/Documents/Testing/GraduationsRequirements2018.pdf

For the 2020 administration year, the paper-based testing was reserved for accommodations only. The computer-based testing was provided via the eMetric-based platform. The online administrations were conducted using the HSA Kiosk web-based software application. The HSA Kiosk allows students to respond to the selected-response (SR) items electronically by selecting an answer choice. Students respond electronically to the constructed-response (CR) items by typing their answers into the response boxes using the computer keyboard. The HSA Kiosk also allows students to respond electronically to the technology-enhanced (TE) items in a variety of formats.

All SR and TE items were machine scored. The CR items were first scored by a human scorer and then received a second score from artificial intelligence (AI) using ACT's Constructed Response Automated Scoring Engine (CRASE+). CRASE+ analyzes a sample of human-scored student responses to produce a model that emulates human scoring behavior. When the scores from the two scorers were adjacent, the higher score was used. When the two scores differed by more than one point, the scoring supervisor would decide on a final resolution score. Additional detailed information about HSA Government and HS MISA is provided below.

HSA Government

The HSA Government exam was administered in January 2020. The May and summer administrations were canceled in 2020 due to COVID-19. The January 2020 administration had two operational item sets and six field test (matrix) item sets. One of the operational item sets was combined with each of three field test item sets. The other operational item set was combined with the other three field test item sets. The result was a total of six distinct test forms for the January 2020 administration.

As just noted, each HSA Government test form consisted of operational and field test items. The operational items were used to produce student scores; students' scores on the field test items were not included in the computation of their scores. For the January administration, field test item performance was analyzed, and all flagged items were reviewed. The field test items that were approved by both the MSDE and Cognia content specialists were then calibrated and marked as available for use in the item bank. Items that were deemed unacceptable were marked as unavailable and may be revised and field tested again in the future. Apart from items selected for public release, which are not reused, the operational items that are returned to the item bank remain unused for at least one year to minimize item exposure.

The operational items in the HSA Government test consisted of SR items, which require students to choose from among four short response options; and brief constructed-response (BCR) items, which require students to write a short response. All items are based on the content outlined in Maryland's Social Studies Standards.[4]

Beginning in 2019, new item types were field-tested as part of the HSA Government test: TE items, including matching, drag and drop, and hot spot items; and evidence-based argument sets (EBAS), which consist of a series of stimuli, SR items, and an extended CR (ECR) item.

Item response models were used to estimate total test scores and subscores via item-pattern scoring. For HSA Government, the three-parameter logistic (3PL) model was used for the SR items (see Section 2 for

[4] The HSA Standards documents can be found on the Maryland School Improvement website at http://www.marylandpublicschools.org/about/Pages/DAAIT/Assessment/HSA/index.aspx

an introduction to item types) and the generalized partial credit model (GPCM) was used for the BCR and ECR items. Refer to *Scale Scores* of Section 4 for the details of the item response theory (IRT) models used and the item-pattern scoring procedure. Total test results on the scale score metric and the performance level based on pass/fail are reported to students. Subscores are not reported to students but are aggregated at the classroom level to provide teachers and administrators with additional information about student performance in each of the subscore categories.

Pre-equated item parameter estimates were used to generate student scores on the Government assessment. When pre-equated item parameter estimates are used, the parameters are not estimated following an administration; instead, existing bank parameter estimates are used to produce student scores. Using this approach, scores can be calculated and assigned to students immediately after their answer documents have been processed.

HS MISA

The HS MISA is the final assessment in a series of science assessments, including the grade 5 and grade 8 MISA, that students take aligned to the NGSS. The HS MISA is given in January and May of each school year. The May 2020 administration of HS MISA was canceled due to COVID-19.

Following the pattern established by the elementary and middle school MISA, the HS MISA consists of item sets that are organized around common stimuli. Students read a stimulus and then answer a set of six questions about the stimulus. These item sets are made up of a combination of multiple selected-response (MSR), SR, TE, and CR items.

The January 2020 HS MISA administration had three operational item sets and nine field test (matrix) item sets. One of the operational item sets was combined with each of three field test item sets. The other operational item sets were combined with the other six field test item sets. The result was a total of nine distinct test forms for the January 2020 administration.

Standard setting for the HS MISA assessment was conducted in August 2019, using a panel of 20 Maryland educators. The panel-recommended cut scores were reviewed by the MSDE. MSDE opted to make small policy-based adjustments to the panel-recommended cut scores. These final cut scores were transformed into scaled scores via the test characteristic curve of the test form used for standard setting. Please see the 2019 HS MISA Standard Setting Report for further details.

Item response models were used to estimate total test scores and subscores via item-pattern scoring. For HS MISA, the two-parameter logistic (2PL) model was used for the SR items and the GPCM was used for non-SR items.

This Maryland HSA technical report consists of eight sections and three appendices.

- Section 1 introduces the Maryland HSA program.
- Section 2 describes the procedures used for test construction and administration.
- Section 3 presents validity evidence for the use of Maryland HSAs.
- Section 4 delineates the scoring procedures and score types.
- Section 5 describes the reporting of 2020 Maryland HSA Government and HS MISA results.
- Section 6 summarizes the results of the analyses of test reliability, decision consistency, and decision accuracy.
- Section 7 provides summary statistics and descriptive information about student characteristics.
- Section 8 gives the results of the analysis of the test data, including classical item analysis, differential item functioning, and field test item calibration and scaling.

- Appendix A provides examples of the score reports.
- Appendix B provides classical item statistics for operational items by administration for both content areas.
- Appendix C provides classical item statistics for field test items by administration for both content areas

Section 2. Test Construction and Administration

Test Development

Planning

For the 2020 High School Assessment Government (HSA Government) test, Cognia content leaders collaborated with their content counterparts at MSDE to build operational forms using selected-response (SR), brief constructed-response (BCR), and technology-enhanced items from the HSA Government item bank. Field test items were embedded in the operational form according to the test design.

For the High School Maryland Integrated Science Assessment (HS MISA), Cognia content leaders collaborated with their content counterparts at MSDE to select operational items according to the test designs. Field test items were selected to continue to build an operational item bank for the HS MISA. In addition, the field test and operational items were planned with consideration to the design of the MISA in grades 5 and 8, to ensure continuity across the science assessments.

In adherence to these considerations, science "clusters" were developed to create a strong, three-dimensional alignment[5] to the Next Generation Science Standards (NGSS), incorporating two NGSS performance expectations. Each cluster was designed around a common stimulus that is based upon valid scientific research and contains six items.

Item Types

As noted in Section 1, four item types were used on the 2020 HSA Government tests:

- SR—questions in multiple-choice format with four answer options and one correct answer;
- BCR—an item type used in Government only, for which the students need to write a short response;
- Technology-enhanced (TE) items—including matching, drag and drop, and hot spot items;
- Evidence-based argument sets (EBAS)—which consist of a series of stimuli, SR items, and an extended constructed-response (ECR) item.

HSA Government

Table 2-1 shows how the operational item types were distributed on each HSA Government form for the 2020 administrations. Each SR item is worth one point, each TE item is worth two points, each BCR is worth four points, and each ECR is worth five points

[5] The Next Generation Science Standards (NGSS) are organized by Performance Expectations (PEs). In the NGSS, the content and the practices of science work together. Therefore, each PE is tied to a Disciplinary Core Idea (DCI) or content piece as well as to a Science and Engineering Practice (SEP) and a Crosscutting Concept (CCC), which are the over-arching science concepts that tie the content and practices. Items developed for Maryland HS Science must be aligned to two, if not all three dimensions of the NGSS.

Table 2-1. Number of Operational Items and Points Possible by Item Type for Each HSA Government Form

	SR	TE	BCR	ECR	Total
Number of items	44	5	2	1	52
Points possible	44	10	8	5	67

HS MISA

As also noted in Section 1, four item types were used on the 2020 HS MISA tests:

- SR—questions in multiple-choice format with four answer options and one correct answer;
- MSR—questions in multiple-choice format with multiple correct answers;
- Constructed-response (CR)—an item type for which the students need to write a response (2-point, 3-point, and 4-point CR items are included on the HS MISA test);
- Technology-enhanced (TE) items—including matching, drag and drop, ordering, graphing, hot spot, fill-in-the-blank (numerical entry only) and inline choice. (1-point and 2-point TE items are included on the HS MISA test).

As previously noted, the operational HS MISA test is designed with item sets, or clusters. Clusters on the operational form contained a stimulus, five machine-scored items (which include SR, MSR, and TE items) and one CR item, in one of three configurations based on the point value of the CR item.

- 2-point CR configuration: three 1-point SR/TE items, two 2-point SR/TE items, one 2-point CR item, or
- 3-point CR configuration: four 1-point SR/TE items, one 2-point SR/TE item, one 3-point CR item, or
- 4-point CR configuration: five 1-point SR/TE items, one 4-point CR item

Table 2-2. Number of Operational Items and Points Possible by Item Type for Each HS MISA Form

	SR, MSR, TE	CR	Total
Number of items	30	6	36
Points possible	36	18	54

Test Specifications and Design

HSA Government

For the HSA Government test, MSDE predetermined the preliminary test design and provided it to Cognia, following the existing HSA Government test blueprints. The final forms were selected by MSDE to adhere to content and psychometric guidelines. The basic test design document provided information based on specified expectations and the distribution of the number of items by item type for each reporting category. The variety of item types represented ensure that a variety of levels of cognitive complexity are addressed, although these levels are not specifically mandated by the test blueprints. Specific items were placed throughout the forms by Cognia content specialists, with the approval of MSDE. Construction of the forms was based on test blueprints approved by MSDE. The HSA Government Operational Blueprint is presented in Table 2-3.

Table 2-3. HSA Government Operational Blueprint

	Total Points Per Category
Standard 1: Civics	32
Standard 2: Peoples of the Nations and World	8
Standard 3: Geography	8
Standard 4: Economic	10
Standard 6: Skills and Processes	9
Total	**67**

Information on the referenced learning goals can be found in the Maryland Social Studies Standards for Government, available on the Maryland School Improvement website at http://www.marylandpublicschools.org/about/Pages/DCAA/Social-Studies/AGHSH.aspx.

HS MISA

For the HS MISA test, MSDE and Cognia worked collaboratively to design an operational form consisting of six NGSS-aligned clusters, each containing one shared stimulus and six items. Each cluster included various item types as outlined above, always including one CR item. The variety of item types represented, as well as the complexity and three-dimensionality of the NGSS ensure that a variety of levels of cognitive complexity are addressed, although these levels are not specifically mandated by the test design.

The HS MISA operational subscore categories and test blueprint are as follows:

- Each test form contained a total of 36 items and 54 possible points, typically in the following cluster configurations: two 2-point CR clusters, two 3-point CR clusters, and two 4-point CR clusters.

- Each test form contained approximately 33 percent Physical Science items, 33 percent Life Science items, and 33 percent Earth and Space Science items across the six operational clusters.
- Each test form contained some same-domain clusters (PS-PS, LS-LS, ESS-ESS) and some integrated clusters (PS-LS, PS-ESS, LS-ESS).

Table 2-4. HS MISA Operational Blueprint

	Approximate Number of Items
Physical Science	12
Life Science	12
Earth and Space Science	12
Total Number of Items	**36**
Total Possible Points	**54**

In addition, test designs are also aligned to groupings of Practices and Crosscutting Concepts as illustrated in Table 2-5.

Table 2-5. Test Design Alignments

Practices Subscore Category	***Min-Max Percentage***		***Crosscutting Concepts Subscore Category***	***Min-Max Percentage***
Investigating and Evaluating (IE) **Investigations* **Data* **Math*	***22-65%*** *(12-35 pts)*		***Patterns and Cause and Effect (PCE)*** **Patterns* **Cause and Effect*	***22-70%*** *(12-38 pts)*
Developing Explanations and Solutions (DES) **Models* **Explanations* **Argument* **Communicating*	***35-78%*** *(19-42 pts)*		***Systems and Their Properties (SP)*** **Scale, Proportion, Quantity* **System and System Models* **Energy and Matter* **Structure and Function* **Stability and Change*	***30-78%*** *(16-42 pts)*

The HS MISA items and clusters were designed to align to a subset of the high school grade band standards, which may be found here: https://mdk12.msde.maryland.gov/Pages/home.aspx. Item development and field test form construction were designed to support future operational test blueprints.

Item Writing

In the 2019-2020 development year, new item development occurred for both the HSA Government and HS MISA tests.

All test items were originally developed by item writers. Item writers were employed to develop high-quality test items that aligned with the Social Studies Standards (Government) or the NGSS. For the HSA Government test, the items were developed by Maryland educators. For HS MISA, item writers were Maryland educators, Cognia content specialists, and Cognia scoring specialists who are experienced in the NGSS. It is anticipated that as the implementation of the NGSS continues, an increasing number of item writers will be Maryland educators.

Item writers were trained on general item writing techniques as well as writing guidelines that are specific to the HS MISA and HSA Government program. After an initial item writer training occurred, follow-up training was provided in the form of individual feedback and specialist review. After this follow-up training occurred, item writers received additional feedback and coaching as necessary.

Upon completion of their writing assignment, the item writers submitted their items to Cognia. Items and clusters that were accepted by the Cognia content team proceeded to the item review and revision process.

Item Review and Revision

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

- Items were edited according to standard rules, including those detailed by the Maryland Overview Document, Style Guide, and Item Specification documents, developed in conjunction with MSDE.
- Items were reviewed for accuracy, organization, comprehension, style, usage, consistency, fairness/sensitivity, and accessibility.
- Item content was reviewed to establish whether the item measured the intended standards.
- Copyright and/or trademark permissions were verified for any materials requiring permissions, for both field test and operational material.
- Items were reviewed by Cognia editorial staff to ensure the item adhered to both the stated MSDE Style Guide and standard grammar rules.
- Internal reviews were conducted, and historical records were established for all version changes.

After Cognia performed the required internal reviews, items were submitted to MSDE for review. MSDE content specialists performed a review of the items and provided feedback to Cognia content specialists. The edits suggested by the MSDE specialists were then incorporated into the items. At this stage, items were also reviewed for accessibility and universal design.

Finally, the items were prepared for review by the Content, Bias/Sensitivity, and Accommodations Review Committees. These committees, selected by MSDE, were composed of diverse groups of Maryland educators. The committees reviewed each item to ensure that the content (a) accurately reflected what was taught in Maryland schools; (b) correctly aligned to the intended standards; (c) did not unfairly favor or disadvantage an individual or group; and (d) was universally designed and accessible to students with disabilities who utilize various presentation and response accommodations.

Upon completion of this final round of reviews, MSDE and Cognia content specialists conducted face-to-face meetings to evaluate and reconcile the reviews. Cognia then applied the requested edits to the items and/or revisions to the accompanying graphics.

For the HSA Government assessment, 191 items were presented for review by the Content, Bias/Sensitivity, and Accommodations Review Committees in 2019. These items were then used to build the 2020 field test forms. Nine items were rejected following committee recommendations and two items were put on hold due to current events or curriculum changes.

For the HS MISA assessment, 36 science clusters were presented for review by the Content, Bias/Sensitivity, and Accommodations Review Committees in 2019. These items were then used to build the 2020 field test forms. These clusters included 36 multi-part stimuli and 540 items. Because of the integrated nature of the clusters, acceptance rates depended on the entire cluster, not individual items. Two clusters were put on hold due to the extent of the revisions requested.

Testing Accommodations

Several alternate test formats were available to test takers, including large-print, braille, and standard paper-based versions of the HSA Government and HS MISA tests. For 2020, all three alternate test formats were available for the January administration in both content areas. For additional information concerning test accommodations see the Maryland Assessment, Accessibility, and Accommodations Policy Manual available here: http://marylandpublicschools.org/programs/Documents/Special-Ed/IEP/MAM508102017.pdf.

Test Construction

HSA Government

The HSA Government forms administered in January of 2020 were constructed using items from the Maryland HSA government item bank. The pool of items that was available for use in the construction of the 2020 forms included items that had been administered, calibrated, and linked to the operational scale. Each HSA Government test form was constructed to meet specific test blueprint specifications. Table 2-2 indicates the distribution of score points associated with each item type.

HS MISA

The HS MISA forms administered in January of 2020 were constructed using items from the 2018 HS MISA stand-alone field tests and the 2019 embedded field test forms. Items flagged for substantial DIF against any of the comparison groups were marked as such in the item bank and they were not used unless required to fulfill content specifications, and then, only after review and approval by MSDE. (See Section 8 for a more detailed account of these analyses and flagging criteria.)

Each HS MISA form was designed to meet the operational test blueprint outlined in tables 2-3 and 2-4 above. Each form was designed with four sessions consisting of two integrated clusters each. Two field test clusters were embedded with the six operational clusters. Each session was designed to be completed in approximately 40 minutes.

As previously stated, each cluster included one shared stimulus and six items. Each cluster contained one CR item worth two, three, or four points. The remaining five items in the cluster were a variety of SR and TE item types.

Item Selection and Form Design

HSA Government

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 this administration, 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 percent of common items across forms was determined by MSDE and is consistent with the test specifications for previous administrations of the HSA Government assessment.

The guidelines used to construct the forms are provided in Table 2-6. The exact composition of the forms varied slightly based on available items in the pool.

Table 2-6. Form Construction Specifications for the HSA Government January 2020 Administration

Form A, B, and C – Operational Core 1	Form AA, AB, AC – Operational Core 2	Form X (Accom.)
Common set ~ 60% Unique items ~ 40%	Common set ~ 60% Unique items ~ 40%	Same as Form A
Field test selection – Unique items	Field test selection – Unique items	Field test selection – Same as Form A

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. Field test items accounted for approximately 19 percent of the total items on each form (12 field test items out of the total of 63 items). The content standards, item types, and item specifications added to the assessment and field tested in 2020 were developed and reviewed by Maryland educators to be representative of the knowledge, concepts, and skills taught in Maryland government courses and designed to be measured by the test.

For this administration, there was more than one form available so the forms were randomly assigned at the student level. Random assignment at the student level means that multiple forms of the test were distributed to students arbitrarily by the computer-based testing platform. Random assignment at the student level helps ensure that all forms are arbitrarily distributed throughout the state.

The 2020 HSA Government 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 Government, the base forms were originally developed in 2003. These base

forms contained BCR items. Between summer 2009 and October 2013, BCR items were discontinued on the HSA Government and the target TCCs for the HSAs were revised so that they were no longer influenced by the characteristics of CR items. Refer to the Educational Testing Service (ETS) memorandum: *Considerations for Setting New Target Test Characteristic Curves for the Maryland High School Assessments (HSAs)* (ETS, 2009) for details on how new target TCCs were created. However, starting in January 2014, BCR items were reintroduced to the HSA Government so the Government target TCCs have been revised back to include BCR items in the calculation of TCCs and CSEMs.

The following general steps were completed during the test construction process for the HSA Government forms:

1. For each administration, all forms were constructed simultaneously in order to provide the best opportunity to construct parallel forms.
2. Items were selected to represent the test blueprint and 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[6] or clanging.[7]
4. After the operational items were selected for the test forms, the field test sets were constructed. Item sets consisted of SR, BCR, TE, and ECR item types. While the field test sets were not constructed to meet any psychometric criteria, they were constructed to meet content criteria. For HSA Government, the field test sets were estimated to be able to be completed by students in approximately 30 to 35 minutes. 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 and 2-2 show the plots of the TCCs and CSEMs of the operational forms used for HSA Government in 2020. The vertical line in each figure represents the proficiency scaled cut score. The CSEMs in Figure 2-2 are CSEM values on the scaled score metric (i.e., scaled CSEMs). HSA Government has only one cut: Proficient. It is important to note that the TCCs and CSEMs shown in the plots are based on pre-equated item parameters and therefore are curves calculated prior to administration of the tests. The TCC plots indicate that all forms for HSA Government were within or very close to each other across the range of scale scores. When 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. When 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 scaled CSEMs were lowest at and above the scaled cut score, which represents the middle and upper ranges of scale scores. Typically, this is where most student scores are located.

[6] ***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.

[7] ***Clanging*** occurs when an identical or similar 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.

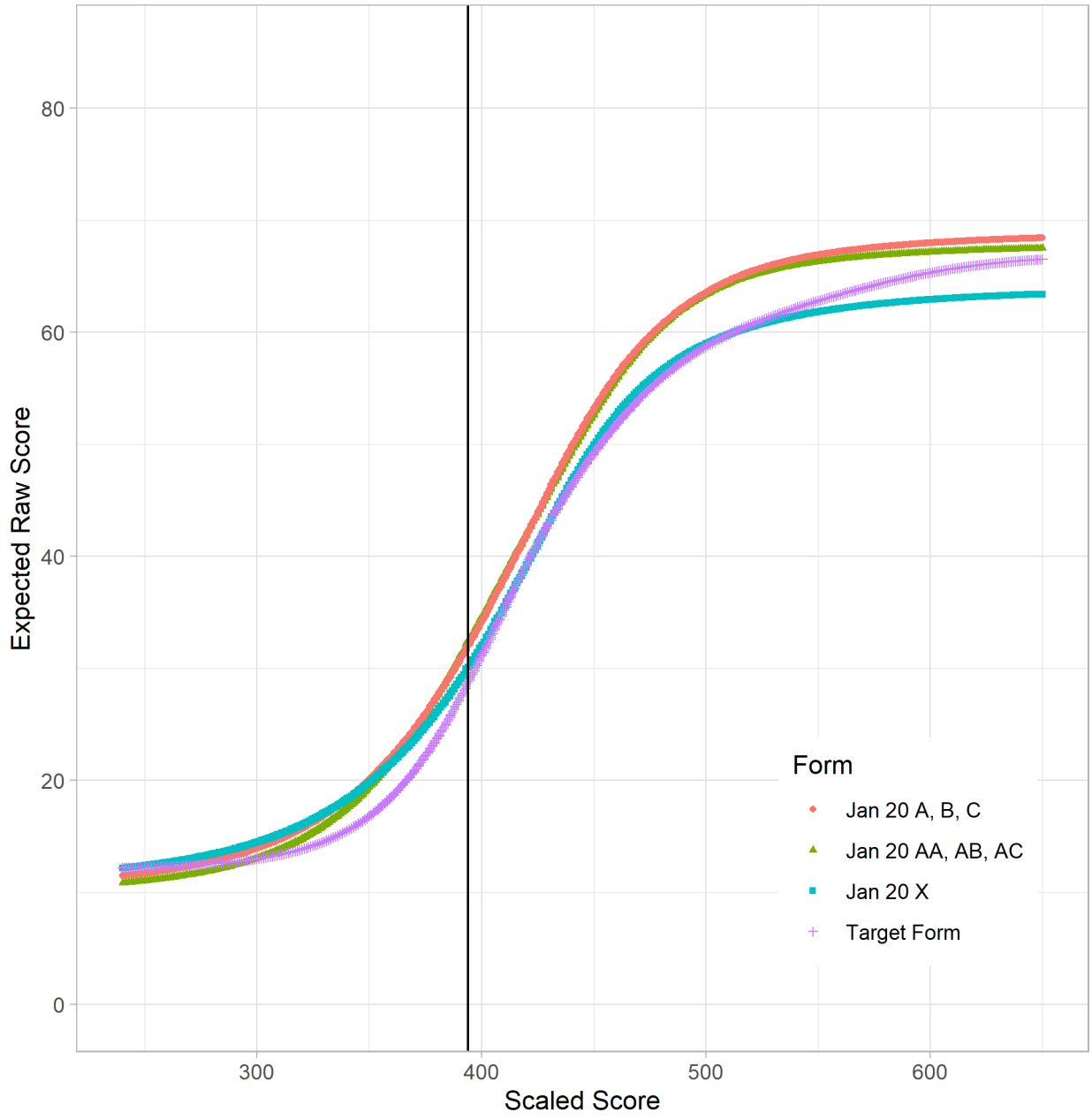


Figure 2-1. Test Characteristic Curves for the 2020 Maryland HSA Government Forms

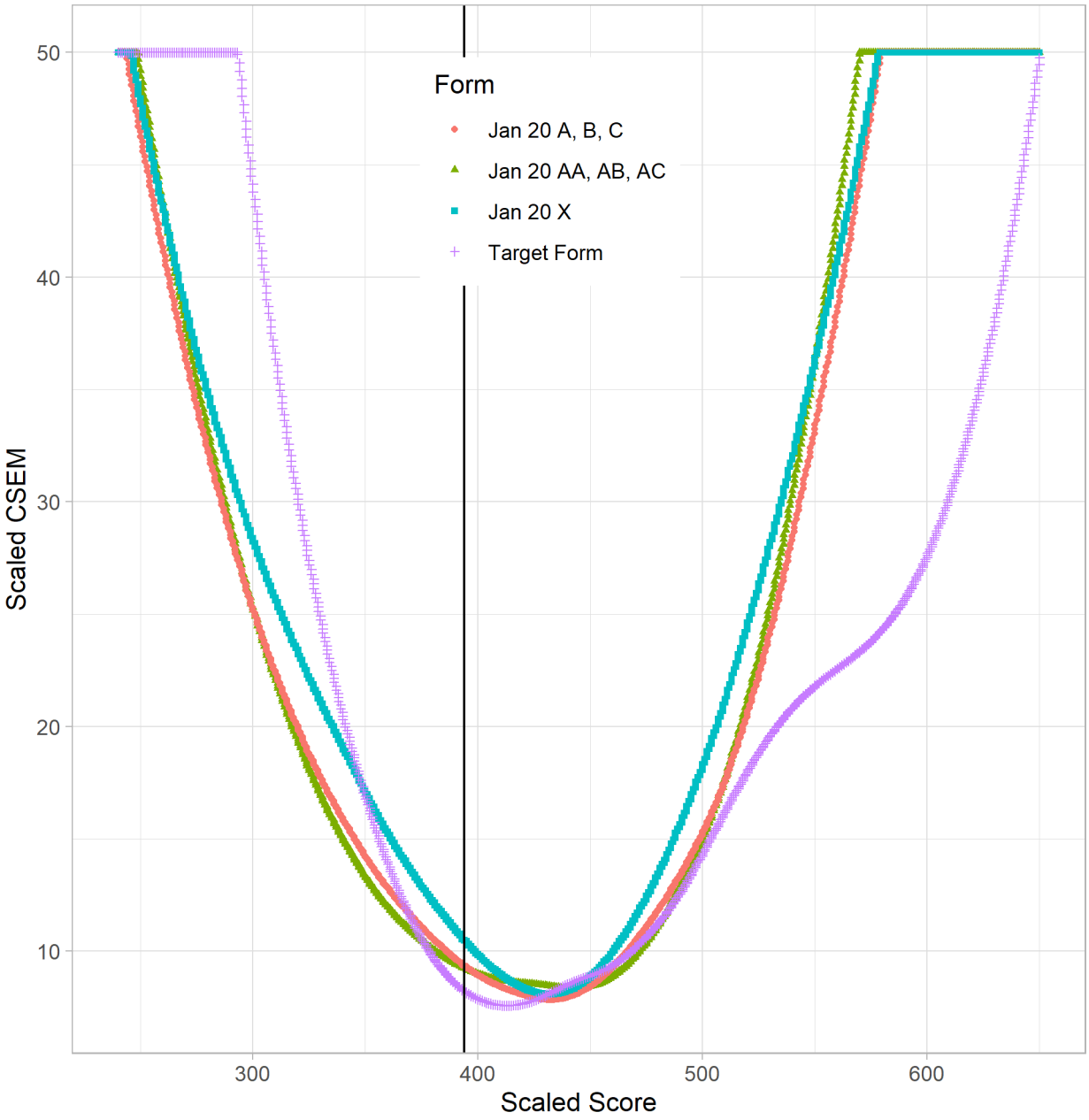


Figure 2-2. Conditional Standard Errors of Measurement and Proficiency Cutoffs for the 2020 Maryland HSA Government Forms

HS MISA

Per the HS MISA test design, when multiple forms were included in an administration, each test form consisted of a common set of operational clusters shared across forms within an administration, as well as a unique set of items. Per this test design, one-half of the operational clusters are shared across the forms for each administration. There were no clusters shared across administrations in 2020, because the May 2020 administration was canceled. However, the clusters that were designed to be shared across administrations were still linked between January 2020 forms.

In addition to the operational items, embedded field test clusters were included with each version of the test form, resulting in multiple versions of a test form containing different sets of field test items. In 2020, six clusters were operational and two were field test clusters.

The guidelines used to construct the forms are provided in Table 2-7. The exact composition of the forms varied slightly based on available items in the pool.

Table 2-7. Form Construction Specifications for the HS MISA January 2020 Administration

Form A, B, C – Operational Core 1	Form AA, AB, AC – Operational Core 2	Form X (Accom.)
Linking clusters – 50% Unique clusters – 50%	Linking clusters – 50% Unique clusters – 50%	Same as Form A
Field test selection – Unique clusters	Field test selection – Unique clusters	Field test selection – Same as Form A

The following general steps were completed during the test construction process for the HS MISA forms:

1. For each administration, all forms were constructed simultaneously in order to provide the best opportunity to construct parallel forms.
2. 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 or clanging.
3. After the operational items were selected for the test forms, the field test sets were constructed. Field test sets consisted of HS MISA clusters across all content areas. While the field test sets were not constructed to meet any psychometric criteria, they were constructed to meet content criteria. 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 continued efforts to build the operational pool of NGSS-aligned HS MISA clusters.

Figures 2-3 and 2-4 show the plots of the TCCs and CSEMs of the forms used for HS MISA in 2020. The vertical lines in each figure represents the scaled cut scores. The CSEMs in Figure 2-4 are CSEM values on the scaled score metric (i.e., scaled CSEMs). HS MISA has three cuts that define four performance levels: Partially Met Expectations, Approach Expectations, Met Expectations, and Exceeded Expectations.

The TCC plots indicate that all forms for HS MISA were within the range of scaled scores, or very close to each other. When 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. When 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 scaled CSEMs were lowest at and above the scaled cut score, which represents the middle and upper ranges of scale scores. Typically, this is where most student scores are located.

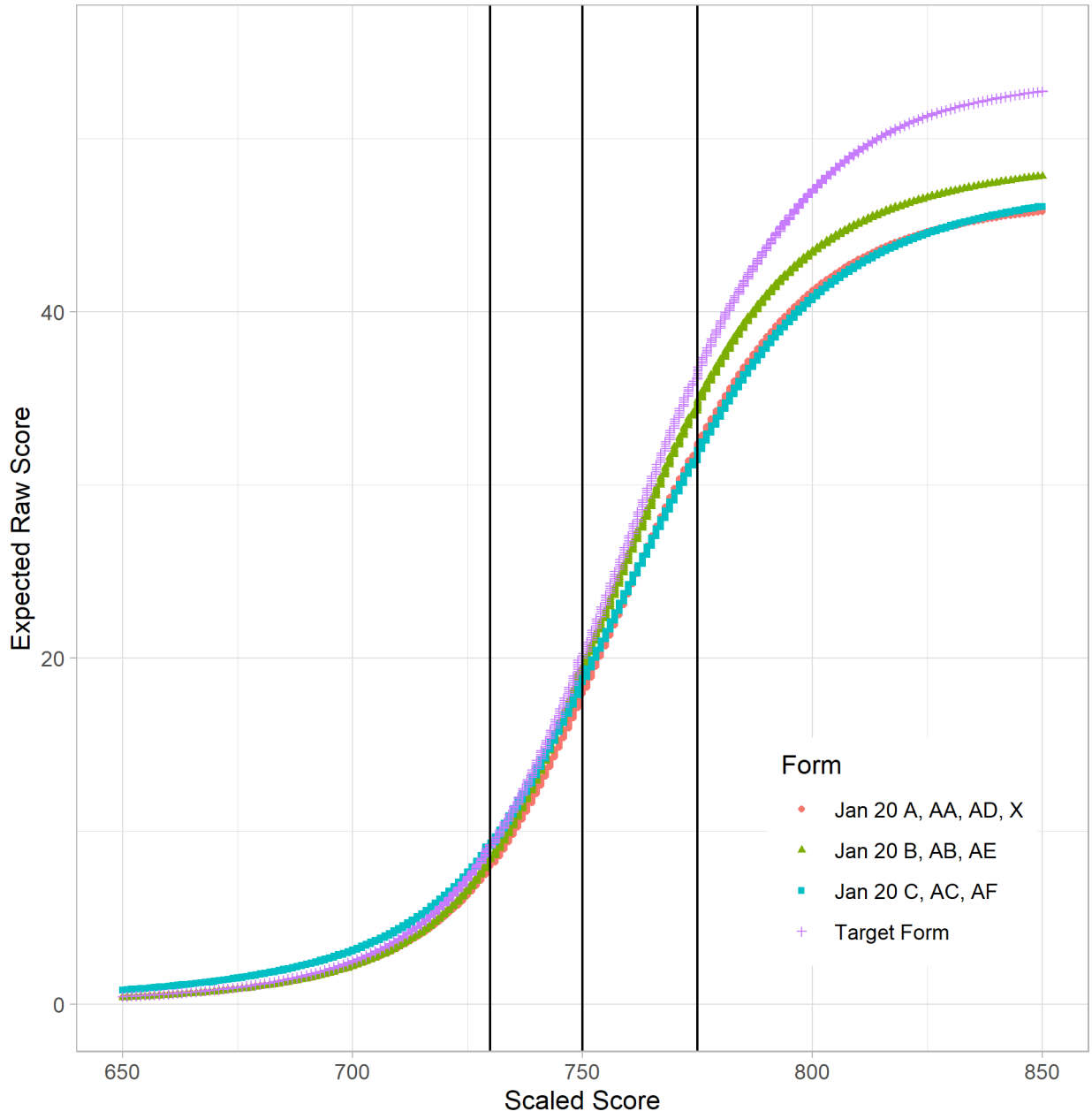


Figure 2-3. Test Characteristic Curves for the 2020 Maryland HS MISA Forms

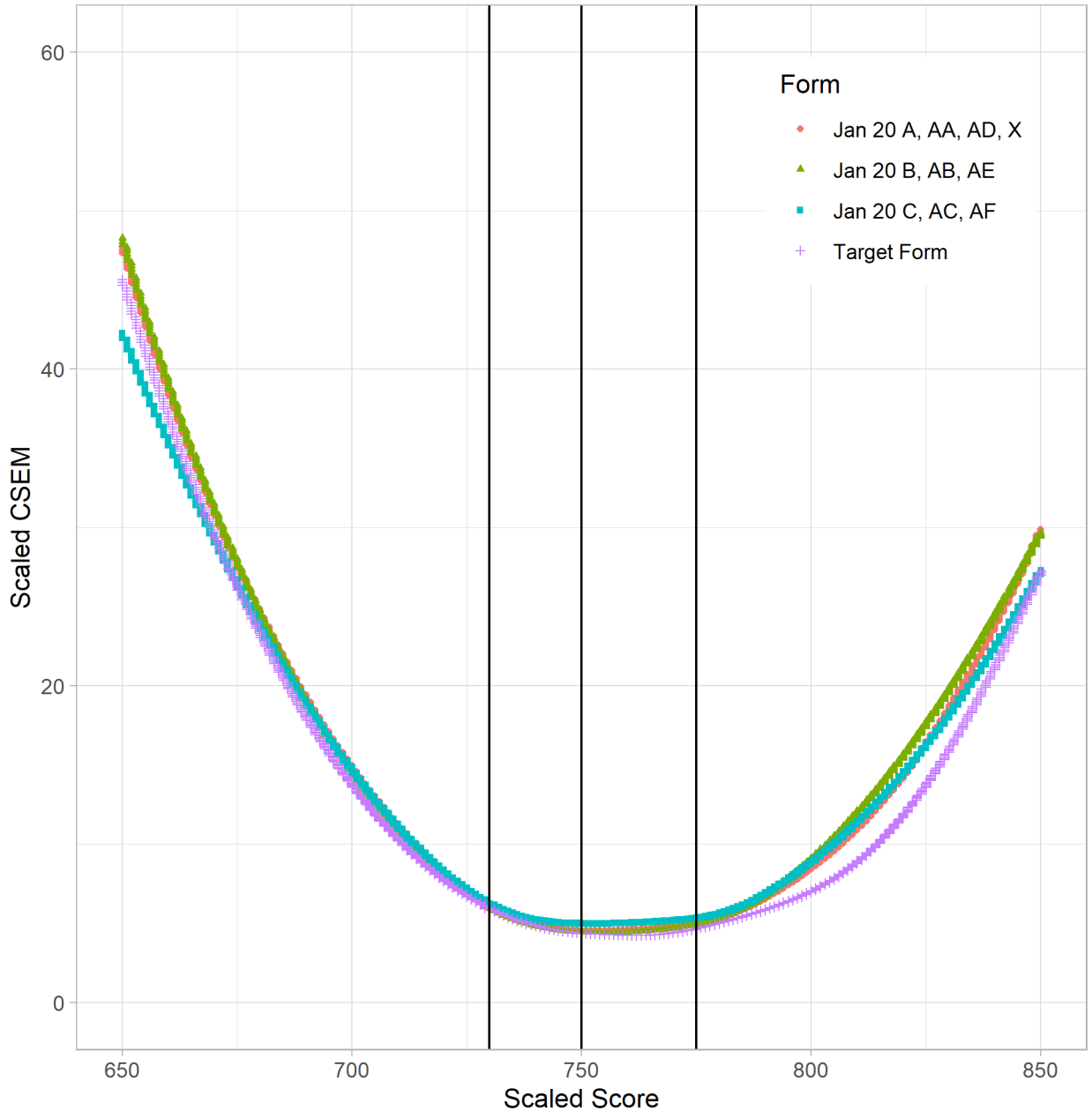


Figure 2-4. Conditional Standard Errors of Measurement and Performance Level Cutoffs for the 2020 Maryland HS MISA Forms

Test Administration

For all Maryland HSA tests administered in 2020, both paper-and-pencil and online versions were available. An online Practice Test was available to the public throughout the administration year.

For all administrations, online forms were randomly assigned. There was one paper form provided for students and used for accommodations or special circumstances. The online and paper test windows were the same durations for the January administration. The online testing window for January was scheduled for a duration of four weeks.

All forms administered without extended time accommodations had timing limits indicated in Table 2-8.

Table 2-8. Test Timing Schedule in Minutes for HSA Government and HS MISA

Content Area	Session One	Break	Session Two	Break	Session Three	Break	Session Four
HS MISA	40 min.	5 min.	40 min.	5 min.	40 min.	5 min.	40 min.
Government	40 min.	5 min.	40 min.	5 min.	40 min.	5 min.	40 min.

Section 3. Validity

Validity is one of the most important attributes of assessment quality and is a fundamental consideration when tests are developed and evaluated (AERA, APA, & NCME, 2014; Messick, 1989). Validity refers to the degree to which logical, empirical, and judgmental evidence supports each proposed interpretation or use of a set of scores. Validity is not based on a single study or type of study but is an ongoing process of gathering evidence to support the interpretation or use of the resulting test scores. The process begins with the test design and continues throughout the entire assessment process, including content specifications, item development, psychometric quality analyses, and inferences made from the test results.

This section provides validity evidence for the High School Assessment Government (HSA Government) and High School Maryland Integrated Science Assessment (HS MISA). Students' scores on the HSA Government and HS MISA are assumed to reflect students' level of knowledge and skills in a content area. The scaled scores on each of these assessments are used to classify students in terms of their level of proficiency based on cut scores established by the state.

Evidence Based on Analyses of Test Content

The HSA Government test is referred to as an end-of-course test because students take it as they complete the appropriate coursework. The HS MISA is the final assessment in a series of science assessments that students take to measure their understanding of the subset of the High School grade band of the Next Generation Science Standards (NGSS). Consequently, HSA Government items are developed to measure the knowledge and skills expected of students following completion of government coursework. The HS MISA items are developed to measure the knowledge and skills expected of students as they complete a variety of high school science courses, because the configuration of high school science courses and the timing of the assessment varies throughout the state. As discussed in Section 2, the development of test content for the HSA Government and the HS MISA is overseen by content experts who have depth of knowledge and teaching experience related to the course(s). Appropriate content leaders who have similar qualifications review the test development work of these individuals.

Evidence based on analyses of test content includes logical analyses that determine the degree to which the items in a test represent the content domain that the test is intended to measure (AERA, APA, & NCME, 2014, p. 14). The test development process for the Maryland HSAs provides numerous opportunities for MSDE to review test content and make changes to ensure that the items measure the knowledge and skills of Maryland students according to course standards. Every item that is created is referenced to a particular instructional standard (goal, expectation, or indicator). During the internal Cognia development process, the specific reference is confirmed or changed to reflect changes to the item. When the item is sent to a committee of Maryland educators for a content review, the members of the committee make independent judgments about the match of the item content to the standard that it is intended to measure and evaluate the appropriateness for intended grade level. These judgments are tabulated and reviewed by the content experts who use the information to decide which items advance to the field test stage of development.

Evidence Based on Analyses of Internal Test Structure

Analyses of the internal structure of a test typically aim to study the relationships among test items and/or test components in order to establish the degree to which the items/components reflect the construct (AERA, APA & NCME, 2014, p. 16). The term "construct" refers to the characteristic that a test is intended to measure and a test score interpretation is based on; in the case of the HSA Government, the construct is the knowledge and skills defined by the test blueprint for each content area.

These test blueprints are derived from the Maryland State Standards for each course. By designing the test blueprints with consideration given to curriculum documents and other expectations for student learning, the blueprints ensure that the content of the test adequately samples the content knowledge and context required for valid inferences about student performance. The test blueprint is presented in Section 2 (see Table 2-2); the Maryland State Standards can be found on the MSDE website at: http://www.marylandpublicschools.org/about/Pages/DAAIT/Assessment/HSA/index.aspx.

High total group internal consistencies as well as similar reliabilities between subgroups with roughly the same sample size provide additional evidence of validity. Measurement error is inevitable. However, high reliability over items within a test implies that the measurement error is small. Coefficient alpha (Cronbach, 1951) and IRT marginal reliability results for each administration for the overall population, as well as for subgroups, can be found in Section 7 of this report in Tables 7-5 through 7-8.

Another way to assess the internal structure of the test is through the evaluation of Pearson correlation matrices between the individual subscores. If subscores are strongly related to each other, it implies a high internal consistency between subscores. Table 3-1 shows the Pearson correlations between subscores of the HSA Government test based on the data from the January administration. Results indicate that each subscore is positively correlated with the overall Scale Score (ranging from 0.69 to 0.80), and that the subscores are positively correlated with each other (ranging from 0.42 to 0.59). It is also noted that the Government subscore correlations are very similar compared to those observed in previous years (e.g., Maryland State Department of Education, 2019).

Table 3-2 shows the Pearson correlations between subscores of the HS MISA test based on the data from the January administration. Results indicate that each subscore is positively correlated with the overall scale score (ranging from 0.78 to 0.96), and that the subscores are positively correlated with each other (ranging from 0.58 to 0.90).

Table 3-1. Correlations Between Subscores—HSA Government

January Administration (N = 18,650)							
	Overall	U.S. Government Structure Functions and Principles	Protecting Rights and Maintaining Order	Systems of Government and U.S. Foreign Policy	Impact of Geography on Governmental Policy	Economic Principles Institutions and Processes	Evaluating Sources Using Evidence Communicating and Critiquing Conclusions
Overall	1.00						
U.S. Government Structure Functions and Principles	0.75	1.00					
Protecting Rights and Maintaining Order	0.80	0.54	1.00				
Systems of Government and U.S. Foreign Policy	0.71	0.51	0.50	1.00			
Impact of Geography on Governmental Policy	0.69	0.51	0.51	0.48	1.00		
Economic Principles Institutions and Processes	0.69	0.45	0.48	0.42	0.43	1.00	
Evaluating Sources Using Evidence Communicating and Critiquing Conclusions	0.75	0.50	0.59	0.46	0.49	0.45	1.00

Table 3-2. Correlations Between Subscores—HS MISA

January Administration (N = 20,289)								
	Overall	Physical Sciences	Life Sciences	Earth and Space Sciences	Investigating and Evaluation	Developing Explanations and Solutions	Patterns and Cause and Effect	Systems and Their Properties
Overall	1.00							
Physical Sciences	0.80	1.00						
Life Sciences	0.84	0.58	1.00					
Earth and Space Sciences	0.84	0.60	0.61	1.00				
Investigating and Evaluation	0.78	0.61	0.66	0.70	1.00			
Developing Explanations and Solutions	0.96	0.75	0.81	0.82	0.65	1.00		
Patterns and Cause and Effect	0.90	0.78	0.81	0.68	0.63	0.90	1.00	
Systems and Their Properties	0.92	0.70	0.75	0.84	0.78	0.86	0.72	1.00

Finally, the internal structures of the HSA Government and HS MISA tests are assessed by the degree to which the test meets the requirements of the statistical models used to estimate item parameters and student scores. Confirmatory factor analysis (CFA) was used to assess the degree to which one-factor models fit the HSA Government and the HS MISA tests. CFA is a useful statistical methodology for evaluating whether performance on items in each test reflects a single underlying characteristic (i.e., a unidimensional test) or a set of distinct characteristics defined by the reporting categories (i.e., a multidimensional test). The CFA results provide evidence as to the degree to which the unidimensional item response theory (IRT) model used to calibrate the HSA Government items is appropriate.

Confirmatory Factor Analyses of the January 2020 Administration Data

To assess the dimensionality of the HSA Government, CFA was conducted using testing data from the January 2020 administration. For HSA Government, Forms A, B, and C contained the same set of operational items, and Forms AA, AB, and AC contained the same set of operational items. Some operational items on Forms A, B, and C were not on Forms AA, AB, and AC. As such, a separate CFA was run per set. CFA was not run on the accommodated form (Form X).

To assess the dimensionality of the HS MISA, CFA was conducted using testing data from the January 2020 administration. For HS MISA, Forms A, AA, and AD contained the same set of operational items, Forms B, AB, and AE contained the same set of operational items, and Forms C, AC, and AF contained the same set of operational items. Some operational items on Forms A, B, and C were not on Forms AA, AB, and AC. None of these three sets of operational items fully overlapped. A separate CFA was run per set. CFA was not run on the accommodated form (Form X).

Mplus (Muthén & Muthén, 2007) was used to calculate matrices of polychoric correlations between the items and was also used to fit specified factor models to the data. In the analysis, the input polychoric correlation matrix was used to estimate the factor loadings between the indicators (items).

Parameters for CFA were estimated using a weighted least-square method with mean and variance adjustment (Muthén, du Toit, & Spisic, 1997). This method leads to a consistent estimator of the model parameters and provides standard errors that are robust under model misspecification. For ordinal data, weighted least squares estimation offers an alternative to full-information maximum likelihood techniques. The latter becomes computationally too demanding for models with more than a few dimensions. Model fit is assessed through a scaled chi-square statistic. However, the degrees of freedom for the reference distribution of this statistic cannot be computed in the standard way. The correct degrees of freedom depend on the data, and hence degrees of freedom may vary when the same model is applied to different data (Muthén, 1998–2004, p. 19-20).

Overall model fit for the CFA model was examined using the scaled chi-square (χ^2) test of model fit in combination with supplemental fit indices. The Tucker-Lewis Index (TLI) compares the chi-square for the hypothesized model with that of the null or "independence" model, in which all correlations or covariances are zero. TLI values range from 0.0 to 1.0; values greater than 0.94 signify good fit (Hu & Bentler, 1999). The comparative fit index (CFI) and root mean square error of approximation (RMSEA) index are both based on non-centrality parameters. The CFI compares the covariance matrix predicted by the model with the observed covariance matrix, and the covariance matrix of the null model with the observed covariance matrix. A CFI value greater than 0.90 indicates acceptable model fit (Hu & Bentler, 1999). The RMSEA assesses the error in the hypothesized model predictions; values less than or equal to 0.06 indicate good fit (Hu & Bentler, 1999).

Table 3-3 shows the results of the analyses. Although the χ^2 statistic was statistically significant ($p <$ .0001), this was expected due to the very large sample size *(N)*. The TLI, CFI, and RMSEA fit statistics indicated that the one-factor solutions generally fit the data well. These fit statistics provide strong evidence in support of the item response theory (IRT) assumption of unidimensionality for both HSA Government and HS MISA.

Table 3-3. Confirmatory Factor Analyses Fit Statistics

Test	Admin.	Forms	# of Factors	# of Items	*N*	*df*	χ^2	*p*-value	TLI	CFI	RMSEA
HSA Govt.	Jan. 2020	Forms A, B, C	1	52	8,336	1,274	6,964.72	< 0.0001	**0.960**	**0.961**	**0.023**
		Forms AA, AB, AC	1	52	8,347	1,274	6,523.60	< 0.0001	**0.963**	**0.964**	**0.022**
HS MISA	Jan. 2020	Forms A, AA, AD	1	36	6,462	594	3,258.11	< 0.0001	**0.979**	**0.980**	**0.026**
		Forms B, AB, AE	1	36	6,480	594	2,939.05	< 0.0001	**0.981**	**0.980**	**0.025**
		Forms C, AC, AF	1	36	6,471	594	2,841.56	< 0.0001	**0.976**	**0.978**	**0.024**

Table entries that meet or exceed the criterion are in bold.

Evidence Based on Response Processes

One source of validity evidence related to response processes is the rate of omitted responses. As part of the validity evidence, the omit rates of the operational items on the HSA Government and HS MISA tests were evaluated. Table 3-4 shows omit rates for operational items from HSA Government and HS MISA by administration and item type.

For both tests, if more than 5 percent of students omit a selected-response (SR) item or more than 15 percent of students omit a non-SR item, that item earns a flag. No operational SR or non-SR items were flagged for HSA Government. For HS MISA, only one non-SR item on the accommodated form was flagged. For that item, 15.2% of the 876 students taking the accommodated form omitted a response. See Appendices A and B for the percentages of students who omitted each item on the HSA Government and the HS MISA test forms.

Other Supporting Information

In addition to the factor analyses and the information regarding speededness presented here and the validation documentation gathered and maintained by MSDE, other information in support of the uses and interpretations of the HSA Government scores appears in the following sections:

- Section 4 provides detailed information concerning the scores that were reported and the cut scores for the HSA Government and HS MISA.
- Section 5 provides detailed information regarding reporting of 2019 Maryland HSA Government and HS MISA results at the student level.

- Section 6 provides information concerning the test characteristics based on classical test theory for the January administration of the HSA Government and HS MISA.
- Section 7 presents information regarding student characteristics for the administration of the HSA Government and HS MISA.
- Section 8 includes documentation regarding the test analyses. Descriptions of classical item analyses and differential item functioning are included. In addition, summary tables of item *p*-value and item-total correlation distributions are provided.

Section 4. Scoring Procedures

Scale Scores

The High School Assessment Government (HSA Government) reporting scale ranges from 240 to 650. For the HSA Government tests, the scale was established in 2003 and defined so that the scale scores had a mean of 400 and a standard deviation of 40.

$$\text{ScaledScore}_{\text{HSA Govt}} = 400 + 40\theta$$

where
θ is the ability level (or pattern score) of a student.

The High School Maryland Integrated Science Assessment (HS MISA) reporting scale ranges from 650 to 850. HS MISA scaled scores are computed via the following:

$$\text{ScaledScore}_{\text{HS MISA}} = 750 + 15.5(\theta - \theta_{Met})$$

where
θ_{Met} is the theta cut score for Met Expectations and is equal to 0.34570.

Students' total test scores and subscores are scale scores derived using item response theory (IRT; Yen & Fitzpatrick, 2006) and item-pattern scoring procedures. HSA Government uses the three-parameter logistic (3PL) model for selected-response (SR) items and the generalized partial credit model (GPCM) for constructed-response (CR) items. HS MISA uses the two-parameter (2PL) model for SR items and the GPCM for non-SR multi-point (polytomous) items.

IRT expresses the probability that a student achieves a certain score on an item (such as correct or incorrect) as a function of the item's statistical properties and the person's ability level (or proficiency level). The 3PL model describes the probability that a person with ability θ responds correctly to item *i* as follows:

$$P_i(\theta) = c_i + (1 - c_i)\frac{\exp[Da_i(\theta - b_i)]}{1 + \exp[Da_i(\theta - b_i)]}$$

where
a_i is the slope parameter of item *i*, characterizing its discrimination;
b_i is the location parameter of item *i*, characterizing its difficulty;
c_i is the lower asymptote parameter of item *i*, reflecting the chance that students with very low proficiency will select the correct answer, sometimes called the "pseudo-guessing" level; and
D is a normal approximation constant.

The 2PL is a special case of the 3PL in which the c-parameter (c_i) is fixed to 0.0.

The GPCM states that the probability that a person with ability θ obtains a score category of *k* on item *i* that has *m* score categories assigned score values ranging from 0 to *m* – 1 can be expressed as:

$$P_{ik}(\theta) = \frac{\exp\left[\sum_{v=1}^{k} a_i(\theta - b_i + d_{iv})\right]}{\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} a_i(\theta - b_i + d_{iv})]}$$

where

b_i is the location parameter for item *i*,

d_{iv} is the step parameter for score *v* on item *i*, and

m is the number of item score categories of item *i* (Muraki, 1992).

An indeterminacy exists in the item parameters of the GPCM. To resolve the indeterminacy, d_0 is fixed to 0 and the sum of the step parameters is fixed to 0.0.

There are essentially two ways of scoring a test: number-correct or item-pattern scoring. Number-correct scoring considers how many test items a student answered correctly in determining that student's total raw score. In contrast, the item-pattern scoring method is based on an IRT model. Item-pattern scoring considers not only a student's total raw responses, but also the psychometric characteristics of test items.

Two students with exactly the same total raw scores will get the same test scores in number-correct scoring. It is highly likely, however, that even though they have the same total raw scores, the actual items they answered correctly were different, and their different sets of correctly answered items could have different item characteristics. In such a case, the students will very likely get different reported test scores in item-pattern scoring. With item-pattern scoring, a student who correctly answers a number of more difficult items will get a higher score than one who answers the same number of easier items. This would be applicable to both total test scores and subscore category scores reported using item-pattern scoring.

Item-pattern scoring has been found to produce smaller standard errors of measurement (SEM) than number-correct scoring. The smaller the SEM, the more confidence we have about the precision of the test results. In addition, test reliability is higher with item-pattern scoring than with number-correct scoring (Yen & Candell, 1991), which means that fewer questions are needed in item-pattern scoring than in number-correct scoring for equivalent scoring accuracy. For these reasons, both total scores and subscores of the HSA Government and HS MISA tests are reported using item-pattern scoring.

Conditional Standard Errors of Measurement

Conditional standard errors of measurement (CSEM) were produced and are equal to the reciprocal of the square root of the test information function (TIF; i.e., the sum of item information functions). CSEMs are standard errors at individual score points, defined as:

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

where

θ is the individual score point (location on the scale),

CSEM(θ*)* is the conditional standard error of measurement at the score point, and

$I(\theta)$ is the test information function value at that score point, θ.

Lowest and Highest Obtainable Test Scores

The maximum likelihood procedure under either the 2PL or 3PL model does not produce finite scale score estimates for students with perfect scores or zero raw scores. In order for all test takers to receive scale scores, scores need to be established for perfect or zero raw scores. Perfect raw scores are assigned the highest obtainable scaled score (HOSS). Zero raw scores are assigned the lowest obtainable scaled score (LOSS). For HSA Government, the LOSS and HOSS are 240 and 650, respectively. For HS MISA, the LOSS and HOSS are 650 and 850, respectively.

Cut Scores

MSDE established the cut scores associated with each of the performance levels in the HSA Government tests in 2003.[8] One cut score, 394, was established for the HSA Government tests in 2003.

MSDE established cut scores for HS MISA in 2019 (Maryland State Department of Education, 2019). HS MISA scaled scores less than 730 fall into the *Partially Met Expectations* performance level. HS MISA scaled scores ranging from 730 to 749 fall into the *Approached Expectations* performance level. HS MISA scaled scores ranging from 750 to 774 fall into the *Met Expectations* performance level. Lastly, HS MISA scaled scores greater than or equal to 750 fall into the *Exceeded Expectations* performance level. More information on HS MISA standard setting can be found in the High School Maryland Integrated Science Assessments (HS-MISA) Standard Setting Report.

Year-to-Year Scale Maintenance

The HSA Government has been pre-equated since 2004. In the pre-equating design, a bank of items with calibrated parameters on the reporting scale must exist before test form construction. The item parameter estimates for new forms are retrieved from the bank and are used to build test forms that are parallel across administrations. Student scores are produced with the existing item parameter estimates; thus scores are linked from one administration to the other.

To expand both the HSA Government and HS MISA item banks, both tests embed field test items in the operational test forms. The field test data for the January administration was calibrated with the operational items at that time. The parameters of field test items were linked to the reporting scale using a fixed item parameter calibration that fixes the item parameters of all operational items to their bank values. Having all operational items serve as linking items ensures that the linking set is large enough to provide stable and reliable results.

[8] 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://archives.marylandpublicschools.org/MsDE/divisions/planningresultstest/Maryland+Standard+Setting+Technical+Reports.htm.

Section 5. Reporting

Reporting of Results

The High School Assessment Government (HSA Government) and High School Maryland Integrated Science Assessment (HS MISA) tests are designed to measure student achievement in the Maryland content standards. Consistent with this purpose, HS Government results are reported in terms of a scaled score and Pass/Fail status. HS MISA results are reported in terms of test scaled scores and performance levels. Performance levels are derived by comparing scaled scores to the scaled cut scores. For HSA Government, there is a single scaled cut score that categorizes student scaled scores into Basic or Proficient. Pass/Fail status on HSA Government is determined by whether a student's scaled score falls at or above the Proficient scaled cut score. For HS MISA, there are three scaled cut scores that categorize student overall scaled scores into the performance levels of Partially Met Expectations, Approached Expectations, Met Expectations, and Exceeded Expectations. Additionally, student MISA integrated dimension performance is reported. Each integrated dimension score is reported as Met or Exceeded Expectations, Approached Expectations, or Partially Met Expectations.

Student results are provided to the Maryland State Department of Education via a secure website. Cognia produces Student Results labels for the HSA Government assessment. Cognia produces the following reports for the HS MISA assessment:

- Student Results Labels
- Individual Student Report
- School Student Roster Report
- School-, District-, and State-Performance Summary Report
- District Summary of Schools Report
- State Summary of Districts Report
- Interactive Reporting

HSA Government Student Results Labels

A Student Results Label is produced for each tested student. Student results labels are printed and mailed to the districts for distribution. The labels provide student identifying information as well as passing scaled score, earned scaled score and pass/fail status for the student.

HS MISA Student Results Labels

A Student Results Label is produced for each tested student. Student results labels are printed and mailed to the districts for distribution. Additionally, labels were available for download via a secure website. The labels provide student identifying information as well as earned scaled score and performance level for the student.

HS MISA Individual Student Results

An Individual Student Results Report is produced for each tested student. Student results reports are printed and mailed to the districts for distribution. Additionally, reports are available for download via a secure website.
The individual student report visualizes the results for the HS MISA assessment which includes the student's overall earned scaled score and performance level. The report also provides a comparison of the

student's performance to the school, district and state as a whole for science. The report provides integrated dimension performance as well (see Appendix A).

HS MISA School Student Roster Report

A School Student Roster Report is produced for each school containing at least one tested student for an administration. Reports are available for download via a secure website. The school student roster report summarizes school, district, and state performance by displaying the average overall scale score as well as the percent of students at each score category for the integrated dimensions. The report provides schools with student performance by listing students' test results.

School-, District-, and State-Performance Summary Report

The Performance Summary Report summarizes HS MISA test results for schools, districts, and the state as a whole as well as by demographic subgroups. The number of valid scores, average scale score, number and percent of students at each performance level statistics are provided for gender, ethnicity/race, economic disadvantage, students with disabilities, and EL demographic subgroups.

District Summary of Schools Report

The District Summary of Schools Report visualizes the HS MISA test results for schools in a particular district. The number of valid scores, average scale score, percent of students at each performance category integrated dimension, percent of students at Science and Engineering Practices and Crosscutting concepts are displayed.

State Summary of Districts Report

The State Summary of District Report visualizes the HS MISA test results for each district. The number of valid scores, average scale score, percent of students at each performance category integrated dimension, percent of students at Science and Engineering Practices and Crosscutting concepts are displayed.

Interactive Reporting

The Performance Level Summary is available in the interactive reporting platform which is a permissions-based Web reporting tool (https://reporting.cognia.org/ReportingMD/login.aspx). To access this report, the user applies basic filtering options, such as the name of the district or school and the grade-level/content-area test. At this point, the user has the option of printing the report for the entire grade level or applying advanced filtering options to select a subgroup of students to analyze. Advanced filtering options include gender, ethnicity, EL, IEP, and FARMS (Free and Reduced Meal Services) A user may provide a custom title for the report for download.

Decision Rules

To ensure that high school assessment results are processed and reported accurately, a document delineating decision rules is prepared before each reporting cycle. The decision rules are observed in the analyses of the high school assessment data and in reporting results. These rules also guide data analysts in identifying students to be excluded from school-, district-, and state-level summary computations.

Quality Assurance

The software quality assurance (SQA) team works together with the data processing and data analysis teams to ensure quality data is captured and delivered accurately. Quality control checks are being performed by the data processors and data analysts as the data is handed off via multiple internal software tools. These quality checks initialize the accuracy of the data being ingested into the database and subsequent tables/columns. The SQA team develops a test plan that includes previously agreed upon report designs and decision rule documents. Test cases housed in an internal test cases repository software are then executed including, but not limited to, the following:

- Testing data counts of data imported.
- Testing data quality of individual fields for valid values, such as gender, ethnicity, etc.
- Validating scripts developed by the software developers to ensure they match business requirements and technical specifications.

Included in this testing effort to ensure the quality of the data, the SQA team uses a sample of schools and districts which is selected based on multiple criteria. A few are identified below.

- Unique student testing records
- Students completed testing
- Students partially completed testing
- Invalidated students

Working together with the data processing and data analysis teams allows for timely and precise turnaround if any data anomalies are found. Test cases are tied to tickets outlining required work to allow for full transparency and cohesive teamwork in validation of the data. Included in the final execution, the SQA team executes test cases validating student printed reports and student labels for accuracy in consistency with the report design specifications. Once all the test cases are passed, the SQA team notifies the Cognia Client Services department for final sign off.

Section 6. Reliability

This section provides the results of test score reliability (classical and IRT-based) and decision consistency and accuracy analyses of the 2020 High School Assessment Government (HSA Government) and High School Maryland Integrated Science Assessment (HS MISA) assessments.

Classical Reliability

The general concept of reliability concerns the precision of a test score. Of interest is quantifying the degree to which a score varies from an average result obtained over many testing occasions due to random factors (Haertel, 2006). A variety of theories and methods can be used to estimate reliability.

Classical test theory defines reliability as the proportion of true-score variance in total score variance. Several different ways of estimating this proportion exist. One commonly used estimate of reliability is coefficient alpha (Cronbach, 1951), an internal consistency measure. It is derived from analysis of the consistency of performance over items within a test and provides a lower-bound estimate of a test's reliability as follows:

$$\alpha \equiv \frac{n}{n-1}\left[1 - \frac{\sum_{i=1}^{n} \sigma_{(Y_i)}^2}{\sigma_x^2}\right]$$

where

n is the number of items,

$\sigma_{(Y_i)}^2$ is the variance of scores on item *i*, and

σ_x^2 is the variance of the total score (sum of scores on the individual items).

Sample estimates are substituted for the population variances in this formula to provide reliability estimates.

IRT Marginal Reliability

IRT marginal reliability estimation is based on applying the standard classical test theory (CTT) formula, relating variances of true score, observed score, and measurement error, in the IRT setting. In CTT, the relationship between these variances is given by:

$$\sigma_X^2 = \sigma_T^2 + \sigma_E^2$$

where

σ_X^2 is the observed-score variance,

σ_T^2 is the true-score variance, and

σ_E^2 is the error variance.

Starting from this basic equation it can be shown that the formula for CTT reliability can be expressed as:

$$CTT\ Reliability = 1 - \frac{\sigma_E^2}{\sigma_X^2}.$$

IRT marginal reliability is based on extending the CTT model to an IRT framework (Samejima, 1994) and provides an IRT-based estimate of the overall test reliability. Error variance is estimated as the mean squared conditional standard error of measurement (CSEM) of the theta estimates across students within a grade. Observed score variance is estimated as the variance of the theta estimates across students within a grade. Equivalently, the mean squared CSEM of the scale scores and the variance of the scale scores can be used in place of the CSEM of the theta estimates and the variance of the theta estimates, respectively. IRT marginal reliability is then given by the following formula:

$$IRT\ Marginal\ Reliability = 1 - \frac{\overline{CSEM(\theta)^2}}{Var(\hat{\theta})} = 1 - \frac{\overline{CSEM(SS)^2}}{Var(SS)},$$

where

$\overline{CSEM(\theta)^2}$ is the mean squared CSEM,

$\overline{CSEM(SS)^2}$ is the mean squared scale CSEM,

$Var(\hat{\theta})$ is the variance of theta estimates, and

$Var(SS)$ is the scale score variance.

Using this formula, IRT marginal reliability estimates were calculated for each multistage test in ELA and mathematics, using the scale scores (and their standard errors) for all the students across all three paths. The reliability of a test can also be evaluated by simply examining directly the CSEMs themselves. CSEMs facilitate the interpretation of individual scale scores. With any given scale score estimate for a student, the reasonable limits of the true scale score for the student can be calculated by using the CSEM for the scale score.

Reliability Results

The total group and subgroup classical and IRT marginal reliabilities are presented in Table 6-1 for HSA Government and Table 6-2 for HS MISA. Note that lower reliability coefficients are sometimes observed when sample sizes are small, the number of repeat test takers is large, or the sample is based only on those taking an accommodated form. That is because under such scenarios, the observed variation in scores tends to be restricted. Such restriction in range can translate to smaller reliability estimates.

Table 6-1. Test Reliability Estimates for HSA Government: January 2020 Forms*

		Forms A–C			Forms AA–AC			Accommodated Form X		
		N	*Alpha*	*IRT*	*N*	*Alpha*	*IRT*	*N*	*Alpha*	*IRT*
Overall		8,336	0.90	0.89	8,347	0.90	0.90	1,967	0.73	0.74
Gender	Male	4,473	0.90	0.90	4,444	0.91	0.90	1,250	0.74	0.74
	Female	3,863	0.90	0.89	3,903	0.90	0.89	717	0.73	0.72
	Missing	0	--	--	0	--	--	0	--	--
Grade	9	757	0.92	0.91	770	0.93	0.91	109	0.83	0.81
	10	2,926	0.92	0.90	2,942	0.92	0.91	538	0.78	0.77
	11	2,826	0.81	0.83	2,859	0.82	0.84	771	0.69	0.70
	12	1,827	0.86	0.86	1,776	0.86	0.87	549	0.70	0.72
	Missing	0	--	--	0	--	--	0	--	--
Special Education	Yes	1,324	0.78	0.84	1,317	0.78	0.84	1,225	0.72	0.74
	No	6,272	0.91	0.89	6,249	0.91	0.90	706	0.70	0.70
	Exited	247	0.88	0.86	263	0.88	0.85	14	--	--
	Exited & placed in 504[a]	51	0.90	0.87	36	--	--	5	--	--
	504	442	0.90	0.88	482	0.90	0.88	17	--	--
Ethnicity	American Indian	28	--	--	14	--	--	2	--	--
	Asian	262	0.93	0.90	233	0.93	0.90	59	0.64	0.59
	African American	3,683	0.82	0.84	3,676	0.83	0.85	731	0.69	0.73
	Hawaiian/ Pacific Islander	13	--	--	12	--	--	1	--	--
	White	2,374	0.92	0.89	2,406	0.92	0.89	395	0.80	0.76
	Hispanic	1,191	0.86	0.88	1,177	0.86	0.89	559	0.69	0.72
	Multi-Ethnic	785	0.88	0.88	828	0.89	0.88	220	0.72	0.72
	Missing	0	--	--	0	--	--	0	--	--
Limited English Proficient	Yes	1,149	0.68	0.81	1,130	0.71	0.84	755	0.67	0.70
	No	6,908	0.91	0.89	6,934	0.91	0.89	1,165	0.76	0.76
	Exited[b]	279	0.80	0.77	283	0.87	0.83	47	--	--

** Statistics not reported for sample size less than 50 (N < 50).*

[a] A 504 plan is a legal document falling under the provisions of the Rehabilitation Act of 1973 that provides a program of instructional services to assist students with special needs who are in a regular education setting.

[b] LEP Exited indicates students who have exited English language acquisition services.

Table 6-2. Test Reliability Estimates for HS MISA: January 2020 Forms*

		Forms A, AA, AD			Forms B, AB, AE			Forms C, AC, AF			Accommodated Form X		
		N	*Alpha*	*IRT*	*N*	*Alpha*	*IRT*	*N*	*Alpha*	*IRT*	*N*	*Alpha*	*IRT*
Overall		6,462	0.90	0.90	6,480	0.89	0.89	6,471	0.88	0.88	876	0.74	0.70
Gender	Male	3,298	0.91	0.91	3,281	0.90	0.90	3,368	0.88	0.88	585	0.74	0.71
	Female	3,164	0.90	0.90	3,199	0.88	0.89	3,103	0.87	0.87	291	0.74	0.69
	Missing	0	--	--	0	--	--	0	--	--	0	--	--
Grade	9	344	0.81	0.87	348	0.81	0.89	319	0.76	0.88	71	0.30	0.72
	10	1,443	0.90	0.90	1,404	0.89	0.89	1,436	0.88	0.88	205	0.71	0.67
	11	4,287	0.90	0.89	4,309	0.89	0.88	4,312	0.87	0.86	487	0.77	0.70
	12	388	0.85	0.86	419	0.85	0.85	404	0.78	0.81	113	0.41	0.52
	Missing	0	--	--	0	--	--	0	--	--	0	--	--
Special Education	Yes	399	0.86	0.86	413	0.83	0.85	434	0.82	0.86	619	0.68	0.64
	No	5,390	0.90	0.90	5,426	0.89	0.89	5,387	0.88	0.87	223	0.37	0.59
	Exited	169	0.89	0.89	188	0.88	0.88	171	0.89	0.88	4	--	--
	Exited & placed in 504[a]	53	0.90	0.89	41	--	--	42	--	--	7	--	--
	504	451	0.89	0.89	412	0.89	0.90	437	0.86	0.86	23	--	--
Ethnicity	American Indian	13	--	--	8	--	--	10	--	--	2	--	--
	Asian	818	0.89	0.89	779	0.88	0.86	767	0.88	0.85	36	--	--
	African American	1,827	0.86	0.87	1,817	0.83	0.86	1,819	0.79	0.84	343	0.50	0.62
	Hawaiian/ Pacific Islander	3	--	--	7	--	--	4	--	--	0	--	--
	White	2,270	0.88	0.87	2,318	0.88	0.86	2,319	0.86	0.85	154	0.86	0.81
	Hispanic	663	0.88	0.88	665	0.87	0.88	660	0.83	0.86	176	0.59	0.51
	Multi-Ethnic	868	0.89	0.89	886	0.87	0.86	892	0.84	0.85	165	0.57	0.56
	Missing	0	--	--	0	--	--	0	--	--	0	--	--
Limited English Proficient	Yes	400	0.61	0.76	385	0.57	0.72	385	0.46	0.75	277	0.29	0.31
	No	5,279	0.90	0.90	5,337	0.89	0.89	5,278	0.88	0.88	524	0.79	0.76
	Exited[b]	783	0.88	0.88	758	0.86	0.84	808	0.82	0.82	75	0.71	0.70

** Statistics not reported for sample size less than 50 (N < 50).*

[a] A 504 plan is a legal document falling under the provisions of the Rehabilitation Act of 1973 that provides a program of instructional services to assist students with special needs who are in a regular education setting.

[b] LEP Exited indicates students who have exited English language acquisition services.

Decision Accuracy and Decision Consistency

For HSA Government tests, students are classified into one of two performance levels: Proficiency or Basic. For HS MISA tests, students are classified into one of four performance levels: Partially Met Expectations, Approached Expectations, Met Expectations, or Exceeded Expectations. The accuracy of decisions based on the specified cut score was assessed for reliability of classification using the computer program called *BB-CLASS* (Brennan, 2004). *BB-CLASS* provides two statistics that describe the reliability of classifications based on test scores (Livingston & Lewis, 1995). Specifically, information from an administration of one form is used to estimate the following:

> ***Decision accuracy***, or the extent to which test takers are classified, on the basis of their estimated ability, into the same performance level as they should be on the basis of their true ability. Decision accuracy addresses the question: How does the actual classification of test takers, based on their single-form scores, agree with the classification that would be made on the basis of their true scores, if their true scores were somehow known?
>
> ***Decision consistency***, or the extent to which test takers are classified into the same performance level if they take the same test one more time. Decision consistency addresses the question: What is the agreement between the classifications based on two non-overlapping, equally difficult forms of the test?

BB-CLASS estimates decision accuracy using an estimated joint distribution of reported performance-level classifications on the current form of the exam and the performance-level classifications based on an all-forms average (true score). *BB-CLASS* estimates decision consistency using an estimated joint distribution of reported performance-level classifications on the current form of the exam and performance-level classifications on the alternate (parallel) form. In each case, the proportion of performance-level classifications with exact agreement is the sum of the entries in the diagonal of the contingency table representing the joint distribution.

Along with the observed frequency distribution of scaled scores, *BB-CLASS* requires an estimate of score reliability for the total test. To that end, IRT marginal reliability was used.

For the January 2020 HSA Government forms, decision accuracy and consistency were calculated across performance levels. The results are provided in Table 6-3. The overall classification estimates are generally high, ranging from .86 to .90. For the January 2020 HS MISA forms, decision accuracy and consistency were also calculated across performance levels. The results are provided in Table 6-4. The overall classification estimates are generally moderate, ranging from 0.68 to 0.81.

Note that in all cases the decision accuracy indices tend to be somewhat larger than the decision consistency indices. This is due to the differences in the estimation procedures. The estimation procedure for decision accuracy includes a random component on one of the two variables, whereas in estimating decision consistency each variable includes a random component (Livingston & Lewis, 1995).

Table 6-3. Decision Accuracy and Consistency: HSA Government January 2020 Forms

Index	Placement Scores	Basic	Proficient	Category Total*
	Forms A–C			
Decision Accuracy	240 – 393	0.63	0.05	0.68
	394 – 650	0.05	0.27	0.32
	Estimated Proportion Correctly Classified*: Total = 0.90			
Decision Consistency	240–393	0.61	0.06	0.67
	394–650	0.08	0.25	0.33
	Estimated Proportion Consistently Classified*: Total = 0.86			
	Form AA–AC			
Decision Accuracy	240–393	0.63	0.04	0.67
	394–650	0.05	0.27	0.33
	Estimated Proportion Correctly Classified*: Total = 0.90			
Decision Consistency	240–393	0.61	0.06	0.67
	394–650	0.07	0.26	0.33
	Estimated Proportion Consistently Classified*: Total = 0.87			
	Accommodated Form X			
Decision Accuracy	240–393	0.93	0.07	1.00
	394–650	0.00	0.00	0.00
	Estimated Proportion Correctly Classified*: Total = 0.93			
Decision Consistency	240–393	0.88	0.06	0.94
	394–650	0.05	0.01	0.06
	Estimated Proportion Consistently Classified*: Total = 0.89			

** Inconsistencies between cell entries and totals are due to rounding.*

Table 6-4. Decision Accuracy and Consistency: HS MISA January 2020 Forms

Index	Placement Scores	Partially Met Expectations	Approached Expectations	Met Expectations	Exceeded Expectations	Category Total*
			Forms –A, AA, AD			
Decision Accuracy	650 - 729	0.11	0.02	0.00	0.00	0.13
	730 - 749	0.03	0.30	0.05	0.00	0.39
	750 - 774	0.00	0.05	0.36	0.02	0.43
	775 - 850	0.00	0.00	0.01	0.04	0.05
	Estimated Proportion Correctly Classified*: Total = 0.81					
Decision Consistency	650 - 729	0.11	0.04	0.00	0.00	0.15
	730 - 749	0.04	0.27	0.07	0.00	0.38
	750 - 774	0.00	0.07	0.33	0.02	0.42
	775 - 850	0.00	0.00	0.02	0.04	0.06
	Estimated Proportion Consistently Classified*: Total = 0.74					
			Forms B, AB, AE			
Decision Accuracy	650 - 729	0.13	0.02	0.00	0.00	0.15
	730 - 749	0.04	0.31	0.05	0.00	0.40
	750 - 774	0.00	0.05	0.35	0.02	0.43
	775 - 850	0.00	0.00	0.01	0.02	0.02
	Estimated Proportion Correctly Classified*: Total = 0.81					
Decision Consistency	650 - 729	0.12	0.04	0.00	0.00	0.16
	730 - 749	0.04	0.28	0.07	0.00	0.39
	750 - 774	0.00	0.07	0.32	0.02	0.41
	775 - 850	0.00	0.00	0.02	0.02	0.04
	Estimated Proportion Consistently Classified*: Total = 0.73					
			Forms C, AC, AF			
Decision Accuracy	650 - 729	0.13	0.03	0.00	0.00	0.15
	730 - 749	0.04	0.29	0.05	0.00	0.38
	750 - 774	0.00	0.06	0.34	0.03	0.44
	775 - 850	0.00	0.00	0.01	0.02	0.03
	Estimated Proportion Correctly Classified*: Total = 0.78					
Decision Consistency	650 - 729	0.12	0.05	0.00	0.00	0.17
	730 - 749	0.04	0.25	0.07	0.00	0.36
	750 - 774	0.00	0.08	0.30	0.03	0.42
	775 - 850	0.00	0.00	0.03	0.03	0.05
	Estimated Proportion Consistently Classified*: Total = 0.70					

continued

Accommodated Form X

Decision Accuracy	650 - 729	0.32	0.08	0.00	0.00	0.41
	730 - 749	0.10	0.41	0.04	0.00	0.55
	750 - 774	0.00	0.01	0.03	0.00	0.04
	775 - 850	0.00	0.00	0.00	0.00	0.00
	Estimated Proportion Correctly Classified*: Total = 0.76					
Decision Consistency	650 - 729	0.29	0.12	0.00	0.00	0.42
	730 - 749	0.12	0.35	0.04	0.00	0.51
	750 - 774	0.00	0.04	0.03	0.00	0.07
	775 - 850	0.00	0.00	0.00	0.00	0.00
	Estimated Proportion Consistently Classified*: Total = 0.68					

** Inconsistencies between cell entries and totals are due to rounding.*

Section 7. Student Characteristics

Summary Statistics

This section presents summary statistics for the January 2020 High School Assessment Government (HSA Government) and High School Maryland Integrated Science Assessment (HS MISA).

Summary statistics (count, mean, and standard deviation) of scale scores in Table 7-1 are reported for all students and by grade for HSA Government and HS MISA. Table 7-2 reports the summary statistics of scores per administration of HSA Government and HS MISA.

Table 7-1. Means and Standard Deviations Overall and by Grade for HSA Government and HS MISA

		N	*Mean*	*SD*
HSA Government				
Overall		18,650	375.9	40.1
Grade				
	9	1,636	395.3	45.6
	10	6,406	388.3	43.2
	11	6,456	366.2	32.1
	12	4,152	364.2	35.6
HS MISA				
Overall		20,289	747.0	17.4
Grade				
	9	1,082	729.7	16.4
	10	4,488	747.7	17.6
	11	13,395	749.3	16.5
	12	1,324	735.7	14.1

Note. Statistics not reported for sample size less than 50 (N < 50). Grade not provided reflects the small number of students whose grade was not provided in the rostering data.

Table 7-2. 2019 Mean Scale Scores by Administration for HSA Government and HS MISA

Content Area	January			May[1]			Summer[1]		
	N	*Mean*	*SD*	*N*	*Mean*	*SD*	*N*	*Mean*	*SD*
HSA Government	18,650	375.9	40.1	--	--	--	--	--	--
HS MISA	20,289	747.0	17.4	--	--	--	--	--	--

[1] *HSA Government and HS MISA were not administered in May and Summer 2020.*

The HSA Government mean scale scores and percentage passing rates are presented for the years 2003 to 2020 in Table 7-3.

Table 7-3. HSA Government Percentage Passing Rates Over Test Years

Year	Mean Scaled Score	Percentage Passing	Percentage Passing – January[1]	Percentage Passing – May[1]	Percentage Passing – Summer[1]
2003	403.5	39.8	--	--	--
2004	406.5	54.6	--	--	--
2005	409.3	67.1	--	--	--
2006	418.5	74.1	--	--	--
2007	417.1	73.3	--	--	--
2008	417.1	71.5	--	--	--
2009	406.3	61.1	--	--	--
2010	408.6	61.7	--	--	--
2011	405.6	62.1	--	--	--
2012	--	*	--	--	--
2013	414.7	72.4	--	--	--
2014	417.6	76.5	--	--	--
2015	412.2	71.8	--	--	--
2016	405.4	62.7	--	--	--
2017	403.6	61.6	--	--	--
2018	403.2	62.5	--	--	--
2019	399.9	60.3	26.4	69.8	29.4
2020[2]	375.9	29.1	29.1	--	--

** The Government test was not administered after the May 2011 administration until January 2013, when it was introduced into the HSAs.*

[1] Prior to 2019, the percent of students passing was not disaggregated by testing window (i.e., January, May, and Summer).

[2] In 2020, HSA Government was only administered in January.

The HS MISA mean scale score and performance level percentage distribution over test years since 2019 are presented in Table 7-4.

Table 7-4. HS MISA Performance Level Percentage Distributions Over Test Windows and Years

Admin/Year	Partially Met Expectations	Approached Expectations	Met Expectations	Exceeded Expectations
January 2019	25.0	42.7	29.8	2.5
May 2019	21.8	43.2	31.3	3.7
January 2020	16.9	38.4	39.4	5.3
May 2020[1]	--	--	--	--
2019-Overall	22.4	43.1	31.0	3.4
2020-Overall[1]	16.9	38.4	39.4	5.3

[1]In 2020, HS MISA was only administered in January.

Summary statistics on HSA Government for all students and for subgroups based on gender, special education programs, ethnicity, and English language proficiency are presented in Table 7-5. Summary statistics on HS MISA for all students and for subgroups based on gender, special education programs, ethnicity, and English language proficiency are presented in Table 7-6. These tables include the numbers of students tested for whom valid scores were available, mean scale scores, and standard deviations of scale scores. In addition, raw score reliabilities are provided for the overall group of test takers and for subgroups. Figure 7-1 shows the distribution of total scale scores for HSA Government for the January

2020 administration. Figure 7-2 shows the distribution of total scale scores for HS MISA for the January 2020 administration.

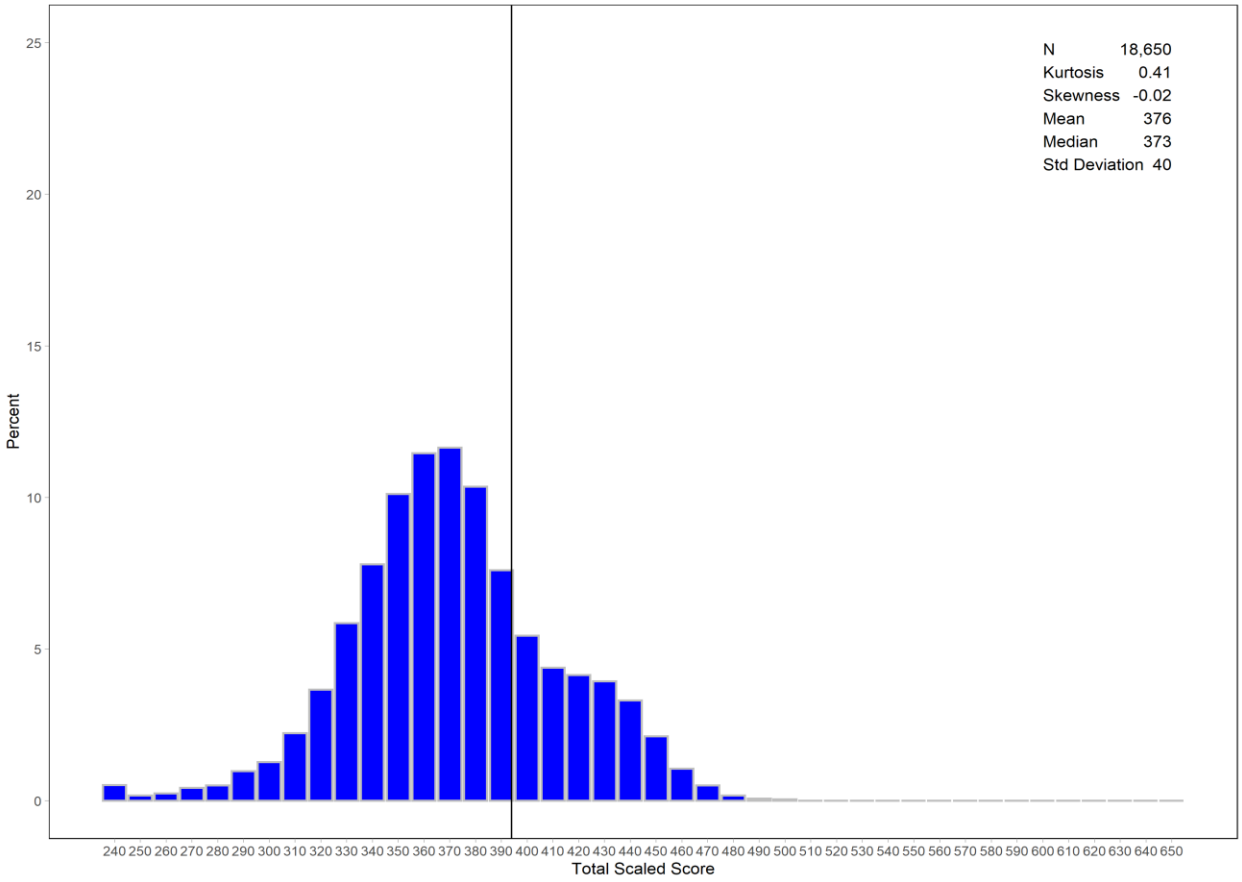


Figure 7-1. Total Scale Score Distribution for HSA Government January 2020 Administration

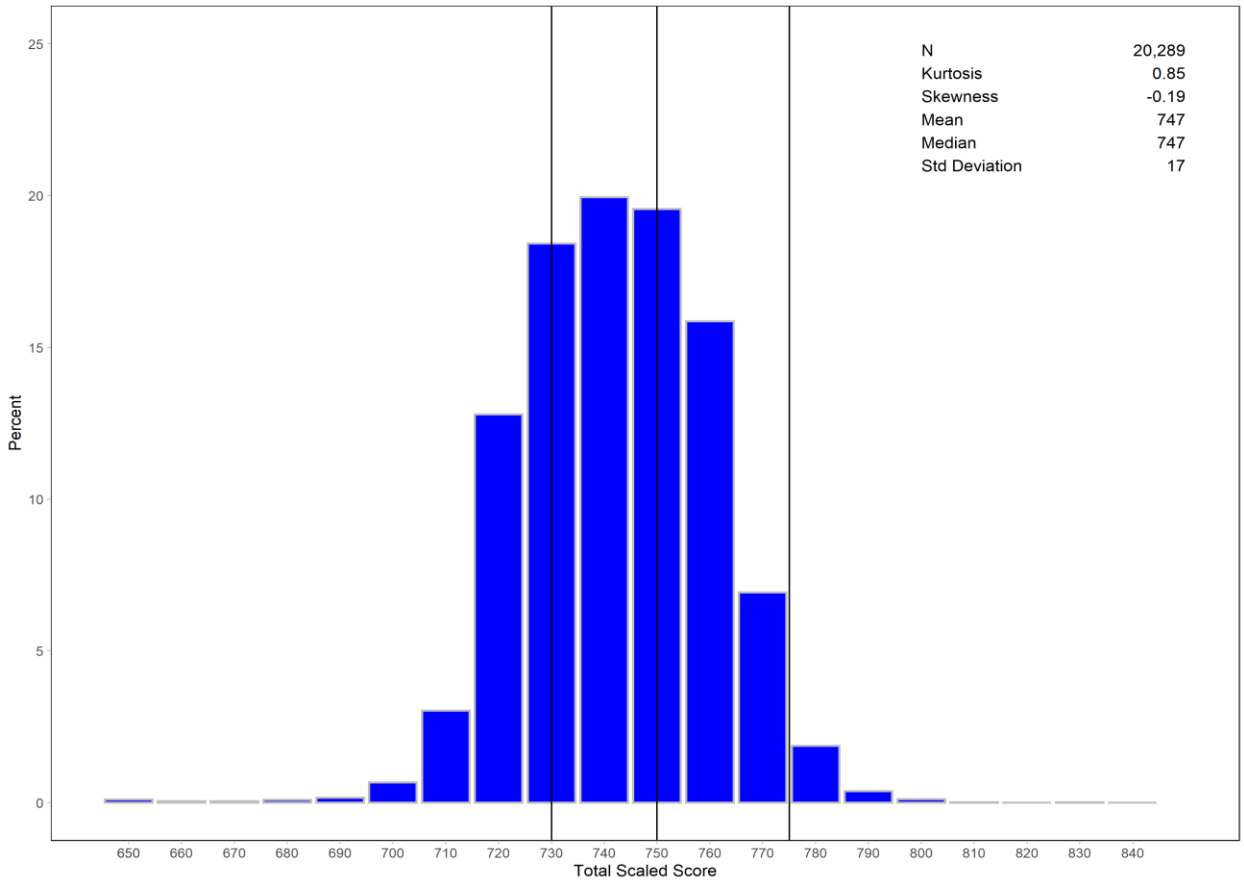


Figure 7-2. Total Scale Score Distribution for HS MISA January 2020 Administration

Table 7-5. Scaled Score Summary Statistics for HSA Government: January 2020 Forms*

		Forms A–C				Forms AA–AC				Accommodated Form X			
		Mean	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*
Overall		378.5	40.3	8,336	100.0	379.1	39.8	8,347	100.0	351.4	31.4	1,967	100.0
Gender	Male	377.7	41.1	4,473	53.7	376.9	41.0	4,444	53.2	350.9	31.9	1,250	63.5
	Female	379.5	39.3	3,863	46.3	381.6	38.4	3,903	46.8	352.2	30.7	717	36.5
	Missing	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0
Grade	9	399.8	43.2	757	9.1	398.7	44.0	770	9.2	339.6	37.0	109	5.5
	10	391.9	42.5	2,926	35.1	391.3	42.4	2,942	35.2	352.7	33.6	538	27.4
	11	367.0	32.2	2,826	33.9	369.1	31.8	2,859	34.3	352.6	29.4	771	39.2
	12	366.2	36.1	1,827	21.9	366.3	35.8	1,776	21.3	350.8	30.4	549	27.9
	Missing	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0
Special Education	Yes	355.9	33.4	1,324	15.9	357.1	32.2	1,317	15.8	350.1	31.7	1,225	62.3
	No	382.7	40.3	6,272	75.2	383.2	40.3	6,249	74.9	351.6	29.5	706	35.9
	Exited	379.3	35.5	247	3.0	379.1	33.3	263	3.2	--	--	14	0.7
	Exited & placed in 504[a]	380.5	37.2	51	0.6	--	--	36	0.4	--	--	5	0.3
	504	386.3	38.5	442	5.3	384.2	37.1	482	5.8	--	--	17	0.9
Ethnicity	American Indian	--	--	28	0.3	--	--	14	0.2	--	--	2	0.1
	Asian	392.1	42.5	262	3.1	396.9	40.3	233	2.8	356.8	25.1	59	3.0
	African American	366.4	33.5	3,683	44.2	367.1	32.8	3,676	44.0	347.1	30.8	731	37.2
	Hawaiian/Pacific Islander	--	--	13	0.2	--	--	12	0.1	--	--	1	0.1
	White	403.8	39.4	2,374	28.5	403.0	39.7	2,406	28.8	360.8	32.8	395	20.1
	Hispanic	364.0	37.7	1,191	14.3	362.8	38.0	1,177	14.1	348.0	30.6	559	28.4
	Multi-Ethnic	376.5	37.8	785	9.4	380.6	36.8	828	9.9	356.2	30.6	220	11.2
	Missing	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0
Limited English Proficient	Yes	351.5	30.3	1,149	13.8	352.2	31.8	1,130	13.5	350.4	29.3	755	38.4
	No	383.0	40.4	6,908	82.9	383.2	39.6	6,934	83.1	351.8	32.7	1,165	59.2
	Exited[b]	379.0	27.6	279	3.3	384.3	31.1	283	3.4	--	--	47	2.4

** Statistics not reported for sample size less than 50 (N < 50).*

[a] A 504 plan is a legal document falling under the provisions of the Rehabilitation Act of 1973 that provides a program of instructional services to assist students with special needs who are in a regular education setting.

[b] LEP Exited indicates students who have exited English language acquisition services.

Table 7-6. Summary Statistics for HS MISA: January 2020 Forms*

		Forms A, AA, AD				Forms B, AB, AE				Forms C, AC, AF				Accommodated Form X			
		Mean	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*	*Mean*	*SD*	*N*	*%*
Overall		748.5	17.2	6,462	100.0	747.0	16.7	6,480	100.0	747.4	18.1	6,471	100.0	732.7	11.4	876	100.0
Gender	Male	748.2	17.7	3,298	51.0	746.5	17.0	3,281	50.6	747.1	18.4	3,368	52.0	732.3	11.5	585	66.8
	Female	748.8	16.7	3,164	49.0	747.6	16.3	3,199	49.4	747.8	17.8	3,103	48.0	733.3	11.2	291	33.2
	Missing	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0
Grade	9	731.2	15.1	344	5.3	729.7	16.6	348	5.4	728.4	18.3	319	4.9	727.7	11.7	71	8.1
	10	748.9	17.2	1,443	22.3	748.0	16.9	1,404	21.7	748.6	18.1	1,436	22.2	730.5	10.9	205	23.4
	11	750.8	16.5	4,287	66.3	749.1	15.6	4,309	66.5	749.5	17.1	4,312	66.6	735.2	11.4	487	55.6
	12	736.8	14.3	388	6.0	736.3	14.0	419	6.5	735.7	14.6	404	6.2	728.8	9.0	113	12.9
	Missing	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0
Special Education	Yes	736.9	14.6	399	6.2	736.0	14.1	413	6.4	735.2	16.8	434	6.7	732.7	10.5	619	70.7
	No	749.2	17.3	5,390	83.4	747.7	16.6	5,426	83.7	748.1	18.0	5,387	83.2	729.5	9.8	223	25.5
	Exited	747.8	15.9	169	2.6	747.4	15.6	188	2.9	749.8	18.0	171	2.6	--	--	4	0.5
	Exited & placed in 504[a]	752.9	16.1	53	0.8	--	--	41	0.6	--	--	42	0.6	--	--	7	0.8
	504	750.1	16.0	451	7.0	748.6	17.0	412	6.4	749.2	16.8	437	6.8	--	--	23	2.6
Ethnicity	American Indian	--	--	13	0.2	--	--	8	0.1	--	--	10	0.2	--	--	2	0.2
	Asian	759.0	16.0	818	12.7	757.8	14.6	779	12.0	758.5	16.3	767	11.9	--	--	36	4.1
	African American	738.9	15.1	1,827	28.3	737.8	14.5	1,817	28.0	738.0	16.0	1,819	28.1	729.6	10.1	343	39.2
	Hawaiian/Pacific Islander	--	--	3	0.0	--	--	7	0.1	--	--	4	0.1	--	--	0	0.0
	White	755.4	14.9	2,270	35.1	753.5	14.7	2,318	35.8	754.2	16.2	2,319	35.8	740.2	14.3	154	17.6
	Hispanic	740.8	15.6	663	10.3	739.2	15.8	665	10.3	739.3	16.8	660	10.2	730.1	8.9	176	20.1
	Multi-Ethnic	746.6	16.0	868	13.4	745.3	14.8	886	13.7	745.3	16.3	892	13.8	733.4	9.5	165	18.8
	Missing	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0	--	--	0	0.0
Limited English Proficient	Yes	730.6	10.9	400	6.2	729.6	10.4	385	5.9	729.6	12.8	385	5.9	730.1	7.5	277	31.6
	No	749.8	17.2	5,279	81.7	748.2	16.7	5,337	82.4	748.7	18.1	5,278	81.6	733.5	12.8	524	59.8
	Exited[b]	748.6	15.2	783	12.1	747.4	13.7	758	11.7	747.5	15.2	808	12.5	736.6	11.5	75	8.6

** Statistics not reported for sample size less than 50 (N < 50).*

[a] A 504 plan is a legal document falling under the provisions of the Rehabilitation Act of 1973 that provides a program of instructional services to assist students with special needs who are in a regular education setting.

[b] LEP Exited indicates students who have exited English language acquisition services.

Demographic Characteristics

Demographic characteristics of the students who took the January 2020 HSA Government and HS MISA tests are presented in Tables 7-7 and 7-8.

Table 7-7. Demographic Information for 2020 HSA Government—Combined Forms

		January		May[c]		Summer[c]	
		N	*%*	*N*	*%*	*N*	*%*
Overall		18,650	100.0	--	--	--	--
Gender	Male	10,167	54.5	--	--	--	--
	Female	8,483	45.5	--	--	--	--
	Missing	0	0.0	--	--	--	--
Special Education	Yes	3,866	20.7	--	--	--	--
	No	13,227	70.9	--	--	--	--
	Exited	524	2.8	--	--	--	--
	Exited & placed in 504[a]	92	0.5	--	--	--	--
	504	941	5.0	--	--	--	--
Ethnicity	American Indian	44	0.2	--	--	--	--
	Asian	554	3.0	--	--	--	--
	African American	8,090	43.4	--	--	--	--
	Hawaiian/ Pacific Islander	26	0.1	--	--	--	--
	White	5,175	27.7	--	--	--	--
	Hispanic	2,927	15.7	--	--	--	--
	Multi-Ethnic	1,833	9.8	--	--	--	--
	Missing	0	0.0	--	--	--	--
Limited English Proficient	Yes	3,034	16.3	--	--	--	--
	No	15,007	80.5	--	--	--	--
	Exited[b]	609	3.3	--	--	--	--

[a] A 504 plan is a legal document falling under the provisions of the Rehabilitation Act of 1973 that provides a program of instructional services to assist students with special needs who are in a regular education setting.

[b] LEP Exited indicates students who have exited English language acquisition services.

[c] In 2020, HSA Government was only administered in January.

Table 7-8. Demographic Information for 2020 HS MISA—Combined Forms

		January		May[c]	
		N	*%*	*N*	*%*
Overall		20,289	100.0	--	--
Gender	Male	10,532	51.9	--	--
	Female	9,757	48.1	--	--
	Missing	0	0.0	--	--
Special Education	Yes	1,865	9.2	--	--
	No	16,426	81.0	--	--
	Exited	532	2.6	--	--
	Exited & placed in 504[a]	143	0.7	--	--
	504	1,323	6.5	--	--
Ethnicity	American Indian	33	0.2	--	--
	Asian	2,400	11.8	--	--
	African American	5,806	28.6	--	--
	Hawaiian/ Pacific Islander	14	0.1	--	--
	White	7,061	34.8	--	--
	Hispanic	2,164	10.7	--	--
	Multi-Ethnic	2,811	13.9	--	--
	Missing	0	0.0	--	--
Limited English Proficient	Yes	1,447	7.1	--	--
	No	16,418	80.9	--	--
	Exited[b]	2,424	11.9	--	--

[a] A 504 plan is a legal document falling under the provisions of the Rehabilitation Act of 1973 that provides a program of instructional services to assist students with special needs who are in a regular education setting.

[b] LEP Exited indicates students who have exited English language acquisition services.

[c] In 2020, HS MISA was only administered in January.

Section 8. Field Test Analyses

Following the receipt of the final score file from eMetric for each administration, analyses were implemented to obtain classical item analyses and differential item functioning (DIF) for High School Assessment Government (HSA Government) and High School Maryland Integrated Science Assessment (HS MISA). Once the classical item analyses were run, the field test items were evaluated psychometrically and submitted to item response theory (IRT) calibration and scaling analyses to obtain IRT item parameter estimates.

Classical Item Analyses

Classical item analyses involve computing a set of statistics based on classical test theory for every item in each form. The statistics provide key information about the quality of the items from an empirical perspective. The following paragraphs outline the statistics estimated for the field test items in the 2020 HSA Government and HS MISA tests. The criteria for flagging the items for content specialists' review are also described below.

Classical item difficulty (p-value): This statistic indicates the mean item score expressed as a proportion of the maximum obtainable item score. For selected-response (SR) items, it is equivalent to the proportion of test takers in the sample that answered the item correctly. For constructed-response (CR) items, the average item score is divided by the maximum score points to obtain the *p*-value. Desired *p*-values for SR items generally fall within the range of 0.25 to 0.90. Occasionally, items that fall outside this range can be justified for inclusion in an item bank based on the quality and educational importance of the item content or the ability to measure students with very high or low achievement, especially if the students have not yet received instruction in the content.

Classical item discrimination (item-total correlation): This statistic describes the relationship between performance on the specific item and performance on the total test, including the item under study. For dichotomously scored items, the item-total correlation is the point-biserial correlation between the key and the total raw score. For polytomously scored items, the item-total correlation is the point-polyserial correlation between the item score and the total raw score. Values less than 0.20 are generally considered to indicate a weaker than desired relationship; therefore, these items receive careful consideration by Cognia and MSDE staff before including them on future forms. Items with negative correlations may indicate serious problems with the item content (e.g., multiple correct answers, incorrect key, unusually complex content, or unfamiliarity with the test content).

Point-biserial correlation of incorrect response option (SR items) with the total raw score: These statistics describe the relationship between selecting an incorrect response option for a specific item and performance on the total test, including the item under study. Typically, the correlation between an incorrect answer and total test performance is weak or negative. Values are typically compared and contrasted with the discrimination index. When the magnitude of a point-biserial correlation for an incorrect answer is strong relative to the correct answer, the item is carefully reviewed for content-related problems. Alternatively, positive point-biserial correlations on incorrect options may indicate that students have not had sufficient opportunity to learn the material.

Percentage of students omitting an item: This statistic is useful for identifying problems with test features, such as testing time and item/test layout. Typically, it is assumed that if students have an adequate amount of testing time, at least 95 percent of them should attempt to answer each question. When a pattern of omit percentages exceeds 5 percent for a series of SR/technology-enhanced (TE) items or 15 percent for CR items at the end of a timed section, this may indicate insufficient time for students to complete all items. For individual items, if the omit percentage is greater than 5 percent for a single SR/TE item or 15 percent for a CR item, this could be an indication of an item/test layout problem. For example, students might accidentally skip an item that follows a lengthy stem.

Proportion of students choosing each response option (SR items): This statistic indicates the proportion of test takers selecting each answer choice, or option. Options not selected by any students or selected by a very low proportion of students may indicate problems with plausibility of the option. Items that do not have all answer options functioning may be discarded or revised and field-tested again.

Frequency distribution of CR score points: Observation of the distribution of scores is useful to identify how well the item is functioning. If no students are assigned the top score point, this may indicate that the item is not functioning with respect to the scoring rubric, there are problems with the item content, or students have not been taught the content.

The following flagging criteria were applied to all field test items administered in 2020:

- *Difficulty flag*: *p*-value is less than 0.10 or greater than 0.90.
- *Discrimination flag:* Item-total correlation is less than 0.10.
- *Distractor flag:* SR point-biserial correlation is positive for an incorrect option, or the magnitude of a point-biserial correlation for an incorrect answer is strong relative to the correct answer.
- *Omit flag:*
    - Percentage omitted is greater than 5 percent for SR or TE items.
    - Percentage omitted is greater than 15 percent for CR items.

Distributions of *p*-values and item-total correlations for the HSA Government field test items administered in January 2020 are presented in Tables 8-1 and 8-2. Corresponding results for the HS MISA field test items administered in January are shown in Tables 8-3 and 8-4. The distribution of *p*-values and item-total correlations in Tables 8-1 to 8-4 are disaggregated between items that are selected-response items and items of all other (non-SR) item types. For both HSA Government and HS MISA, the non-SR item types were TE, MSR, and CR.

The corresponding item-level classical statistics are presented in Appendix B.

Table 8-1. Distribution of *p*-Values for HSA Government January 2020 Field Test Items

	SR Items		Non-SR Items	
	N	%	N	%
$p < 0.10$	0	0	0	0
$0.10 \leq p < 0.20$	0	0	3	12
$0.20 < p < 0.30$	2	7	4	16
$0.30 \leq p < 0.40$	8	27	4	16
$0.40 \leq p < 0.50$	9	30	1	4
$0.50 \leq p < 0.60$	4	13	7	28
$0.60 \leq p < 0.70$	6	20	4	16
$0.70 \leq p < 0.80$	1	3	1	4
$0.80 \leq p < 0.90$	0	0	1	4
$p \geq 0.90$	0	0	0	0
Descriptive Statistics				
Number of Items	30		25	
Mean	0.48		0.45	
SD	0.14		0.19	
Min	0.23		0.15	
Max	0.75		0.81	

Note. SR = Selected-Response Items, Non-SR items include all item types other than SR items.

Table 8-2. Distribution of Item-Total Correlations for HSA Government January 2020 Field Test Items

	SR Items		Non-SR Items	
	N	%	N	%
$r < 0.10$	1	3	0	0
$0.10 \leq r < 0.20$	2	7	0	0
$0.20 < r < 0.30$	8	27	4	16
$0.30 \leq r < 0.40$	7	23	6	24
$0.40 \leq r < 0.50$	10	33	6	24
$0.50 \leq r < 0.60$	2	7	2	8
$0.60 \leq r < 0.70$	0	0	2	8
$0.70 \leq r < 0.80$	0	0	5	20
$0.80 \leq r < 0.90$	0	0	0	0
$r \geq 0.90$	0	0	0	0
Descriptive Statistics				
Number of Items	30		25	
Mean	0.35		0.48	
SD	0.12		0.18	
Min	0.03		0.21	
Max	0.51		0.75	

Note. SR = Selected-Response Items, Non-SR items include all item types other than SR items.

Table 8-3. Distribution of *p*-Values for HS MISA January 2020 Field Test Items

	SR Items		Non-SR Items	
	N	%	N	%
$p < 0.10$	0	0	0	0
$0.10 \leq p < 0.20$	0	0	9	29
$0.20 < p < 0.30$	4	10	8	26
$0.30 \leq p < 0.40$	12	29	4	13
$0.40 \leq p < 0.50$	11	26	6	19
$0.50 \leq p < 0.60$	12	29	2	6
$0.60 \leq p < 0.70$	2	5	2	6
$0.70 \leq p < 0.80$	1	2	0	0
$0.80 \leq p < 0.90$	0	0	0	0
$p \geq 0.90$	0	0	0	0
Descriptive Statistics				
Number of Items	42		31	
Mean	0.45		0.32	
SD	0.11		0.16	
Min	0.25		0.10	
Max	0.76		0.65	

Note. SR = Selected-Response Items, Non-SR items include all item types other than SR items.

Table 8-4. Distribution of Item-Total Correlations for HS MISA January 2020 Field Test Items

	SR Items		Non-SR Items	
	N	%	N	%
$r < 0.10$	3	7	1	3
$0.10 \leq r < 0.20$	9	21	2	6
$0.20 < r < 0.30$	7	17	7	23
$0.30 \leq r < 0.40$	10	24	5	16
$0.40 \leq r < 0.50$	9	21	7	23
$0.50 \leq r < 0.60$	4	10	7	23
$0.60 \leq r < 0.70$	0	0	2	6
$0.70 \leq r < 0.80$	0	0	0	0
$0.80 \leq r < 0.90$	0	0	0	0
$r \geq 0.90$	0	0	0	0
Descriptive Statistics				
Number of Items	42		31	
Mean	0.31		0.39	
SD	0.16		0.15	
Min	-0.11		0.05	
Max	0.56		0.67	

Note. SR = Selected-Response Items, Non-SR items include all item types other than SR items.

Differential Item Functioning

Following the classical item analyses, differential item functioning (DIF) analyses were performed for HSA Government and HS MISA. One goal of test development is to assemble a set of items that provides an estimate of student ability that is as fair and accurate as possible for all groups within the population. DIF statistics are used to identify items in which focal groups of students (e.g., Females, African Americans, Hispanics) with the same underlying level of ability have different probabilities than reference groups (e.g., Males, Whites) of answering correctly. If the item is more difficult or easier for an identifiable focal subgroup, the item may be measuring something different than the intended construct. However, it is important to recognize that DIF-flagged items might be related to actual differences in relevant knowledge or skill (item impact) or statistical Type I error. A subsequent review by MSDE and Cognia content experts was conducted to investigate the source and meaning of evident differences.

The following groups were included in DIF comparison:

- Females (focal)—Males (reference)
- African Americans (focal)—Whites (reference)
- Hispanics (focal)—Whites (reference)
- Asian (focal)—Whites (reference)
- Hawaiian/Pacific Islander (focal)—Whites (reference)
- American Indian/Alaska Native (focal)—Whites (reference)
- English Language Learner (ELL) (focal)—Non-ELL (reference)
- Special Education (focal)—Non-Special Education (reference)

Cognia used the standardization method for dichotomous and polytomous items (Dorans & Kulick, 1986).

The standardization procedure (Dorans & Kulick, 1986; Dorans & Holland, 1993) is used in conjunction with the Mantel chi-square statistic (e.g., Holland & Thayer, 1988). In the standardization method, the matching variable is the total score on all items and the differences in the item score between the two comparison groups are calculated for each item. The standardized mean difference for the item is the weighted average of these differences, where the relative frequency of the focal group at each score point serves as the weighting function.

The flagging criteria for DIF are listed below. Positive values favor the focal group and negative values favor the reference group. The same DIF flagging criteria are used for HSA Government and HS MISA.

A) The item is classified as negligible DIF (A), if the Mantel Chi-square p-value ≤ 0.05; or the Mantel Chi-square p-value < 0.05 and the Standardized Mean Difference $|SMD/SD| \leq 0.17$.
B) The item is classified as moderate DIF (B), if the Mantel Chi-square p-value < 0.05 and $|SMD/SD|$ is between 0.17 and 0.25.
C) The item is classified as severe DIF (C), if the Mantel Chi-square p-value < 0.05 and $|SMD/SD| > 0.25$.

IRT Calibration and Scaling

In terms of operational items, the January 2020 forms of HSA Government and HS MISA were pre-equated. Test scoring was performed via IRT pattern scoring on the set of operational items on a given test form. The IRT parameters for the operational items were fixed to their item bank values. As noted in Section 4, the IRT models used to calibrate the HSA Government field test items are the 3-parameter logistic (3PL) model for SR items and the generalized partial credit model (GPCM) for CR items. The

IRT models used to calibrate the HS MISA field test items are the 2-parameter logistic (2PL) model for SR items and the GPCM for non-SR items.

In terms of field test items, following the classical item analyses, the field test items from the HSA Government and HS MISA January administration were evaluated and then submitted to IRT calibration and scaling.

Before calibration, the items with poor classical item statistics and the items that were not scored per MSDE's instructions were removed (see Figure 8-1). These items have been identified for revision and possible additional field testing. The items excluded from HSA Government and HS MISA calibrations are listed in Tables 8-5 and 8-6, respectively.

Table 8-5. Maryland HSA Government Field Test Items Excluded from Calibration

Admin.	ItemID	Form(s)	Response Type	Reason
January	006QOO	B20	MC	Low item-total correlation (r = 0.03).

Table 8-6. Maryland HS MISA Field Test Items Excluded from Calibration

Admin.	ItemID	Form(s)	Response Type	Reason
January	007P6B	A20, AA20	MC	Near-zero item-total correlation.
January	006MA5	B20	MC	Negative item-total correlation.

Tables 8-7 and 8-8 present the number of field test items that were flagged for further review and evaluation, for HSA Government and HS MISA, respectively.

Table 8-7. HSA Government Field Test Items Flagged for Further Review

Admin.	*p*-Value <0.10	*p*-Value >0.90	Item-Total Corr <0.10	Distractor Item-Total Corr>0	Omit Rate[c]	C-level DIF	Missing Response[a]	Total Flagged	# of Items[b]
January 2020	0	0	0	4	0	1	0	5	54
May 2020[d]	--	--	--	--	--	--	--	--	--

[a] Responded by 0 students; [b] Represents total number of unique items; [c]5% for MC items and 15% for non-MC items. [d]In 2020, HSA Government was only administered in January.

Table 8-8. HS MISA Field Test Items with Statistical Flags Retained in Calibration

Admin.	*p*-Value <0.10	*p*-Value >0.90	Item-Total Corr <0.10	Distractor Item-Total Corr>0	Omit Rate[c]	C-level DIF	Missing Response[a]	Total Flagged	# of Items[b]
January 2020	0	0	2	9	0	1	0	12	73
May 2020	--	--	--	--	--	--	--	--	--

[a] Responded by 0 students; [b] Represents total number of unique items; [c]5% for MC items and 15% for non-MC items. [d]In 2020, HS MISA was only administered in January.

The computer program PARSCALE 4.1 (Muraki & Bock, 2003) was used for all item calibration. PARSCALE is a well-recognized IRT calibration software in the industry, and it is capable of calibrating items with both dichotomous and polytomous data using a variety of dichotomous and polytomous IRT

models. Because it is specifically designed for IRT calibration, it is fast and efficient. The calibration and equating process is outlined in the steps below.

1. For each test, a scored item response matrix with a sparse design is assembled. Essentially, this means that the data were set up using the format presented in Figure 8-1. In the figure, Xs represent items, while spaces indicate missing data. For example, items included on version 2 but not on version 1, 3, 4, or 5 were treated as "not administered" for the purposes of the analyses and are denoted as "missing" in the figure.

Common	Unique 1	Unique 2	Unique 3	Unique 4	Unique 5
XXXXXXXX	XXXXXXXX				
XXXXXXXX		XXXXXXXX			
XXXXXXXX			XXXXXXXX		
XXXXXXXX				XXXXXXXX	
XXXXXXXX					XXXXXXXX
Common	Unique 1				
Common		Unique 2			
Common			Unique 3		
Common				Unique 4	
Common					Unique 5

Figure 8-1. Sparse Matrix Design for Field Test Item Calibration

2. All items are calibrated, and the results were reviewed to determine if any items failed to calibrate appropriately.

In the final calibration, the item parameters for the field test items are freely estimated, with the item parameters for all operational items fixed to their bank values. This means the operational items place the field test items onto the operational reporting scale. Once the items were calibrated and placed onto the operational scale, they were loaded into the item bank.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Brennan, R. L. (2004). *Manual for BB-CLASS: A Computer Program that Uses the Beta-Binomial Model for Classification Consistency and Accuracy Version 1.1*. University of Iowa, Iowa City, IA.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*, 292–334.

Dorans, N. J., & Holland, P. W. (1993). DIF detection and description. In P. W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates.

Dorans, N. J., & Kulick, E. (1986). Demonstrating the utility of the standardization approach to assessing unexpected differential item performance on the Scholastic Aptitude Test. *Journal of Educational Measurement 23*, 355–368.

Educational Testing Service (2009). *Considerations for setting new target test characteristic curves for the Maryland High School Assessments*. Princeton, NJ: Educational Testing Service.

Educational Testing Service (October 29, 2010). *Study of the comparability of online and paper forms of the May 2010 Maryland High School Assessments*. Princeton, NJ: Educational Testing Service.

Haertel, E. H. (2006). Reliability. In R. L. Brennan (Ed.) *Educational measurement* (4th edition, pp. 64–110). Westport, CT: Praeger.

Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–145). Hillsdale, NJ: Erlbaum.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6(1),* 1–55.

Livingston, S. A., & Lewis, C. (1995). Estimating the consistency and accuracy of classification based on test scores. *Journal of Educational Measurement, 32*, 179–197.

Maryland State Department of Education. (2019). High School Maryland Integrated Science Assessments (HS-MISA) Standard Setting Report. Baltimore, MD: Maryland State Department of Education.

Messick, S. (1989). Validity. In R.L. Linn (Ed.) *Educational measurement* (3rd edition, pp. 13–103). New York, NY: American Council on Education and Macmillan.

Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *ETS Research Report, RR-92-6*. Princeton, NJ: Educational Testing Service.

Muraki, E., & Bock, R. D. (2003). *PARSCALE 4.1 [Computer software]*. Lincolnwood, IL: Scientific Software International.

Muthén, B.O. (1998–2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.

Muthén, B., du Toit, S.H.C., & Spisic, D. (1997*). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes*. Unpublished manuscript, University of California, Los Angeles.

Muthén B. O., & Muthén, L. K. (2007). *Mplus 5* [Computer program]. Los Angeles, CA: Muthén & Muthén.

Yen W. M., Candell G. L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education, 4*, 209–228.

Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed*.), Educational measurement* (4th ed., pp. 111–153). Westport, CT: American Council on Education & Praeger.

Appendix A. Score Reports

MCAP
Maryland Comprehensive Assessment Program

DISTRICT SUMMARY OF SCHOOLS

CONFIDENTIAL - DO NOT DISTRIBUTE

Demonstration District A
MARYLAND

HIGH SCHOOL SCIENCE ASSESSMENT, 2019-2020

JANUARY 2020

PERFORMANCE DISTRIBUTION BY %	NUMBER OF VALID SCORES	AVERAGE SCALE SCORE	INTEGRATED DIMENSIONS			SCIENCE and ENGINEERING PRACTICES (SEP)*		CROSSCUTTING CONCEPTS (CCC)*	
			Physical Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts*	Life Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts*	Earth and Space Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts*	Investigating and Evaluation (IE)	Developing Explanations and Solutions (DES)	Patterns and Cause and Effect (PCE)	Systems and Their Properties (SP)
STATE	20,289	747	11 27 62	8 24 68	8 30 62	32 34 33	34 33 34	33 33 34	33 32 35
17 38 39 5									
DISTRICT	163	743	10 37 53	6 28 66	7 39 55	37 38 25	34 48 17	34 52 15	36 41 23
13 53 33 1									
Demonstration School 1	6	738	17 50 33	17 50 33	17 33 50	33 67 0	50 33 17	67 33 0	50 33 17
33 50 17 0									
Demonstration School 2	157	743	10 36 54	6 27 67	6 39 55	37 37 26	34 49 17	32 52 15	35 41 24
13 53 34 1									

2 Partially Met Expectations (650-729) | 3 Approached Expectations (730-749) | 4 Met Expectations (750-774) | 5 Exceeded Expectations (775-850)

Partially Met Expectations | Approached Expectations | Met or Exceeded Expectations

Lower | Middle | Upper

This report is NOT for public review. Distribution within your school/district must be in accordance with state and federal privacy laws, and local school board policy.
* Numbers are percentages

03/23/2020

DISTRICT PERFORMANCE LEVEL SUMMARY REPORT

MCAP
Maryland Comprehensive Assessment Program

CONFIDENTIAL - DO NOT DISTRIBUTE

Demonstration District A
MARYLAND

JANUARY 2020

HIGH SCHOOL SCIENCE ASSESSMENT, 2019-2020

Purpose: This report describes group achievement in terms of average scale scores and performance levels.

	Number of Valid Scores	Average Scale Score	Performance Levels								≥ Level 4 Met or Exceeded Expectations	
			Level 2 Partially Met Expectations		Level 3 Approached Expectations		Level 4 Met Expectations		Level 5 Exceeded Expectations			
			#	%	#	%	#	%	#	%	#	%
State	**20,289**	**747**	**3,435**	**17**	**7,785**	**38**	**7,999**	**39**	**1,070**	**5**	**9,069**	**45**
District	163	743	22	13	86	53	54	33	1	1	55	34
Gender												
Female	69	744	9	13	29	42	31	45	0	0	31	45
Male	94	742	13	14	57	61	23	24	1	1	24	26
Ethnicity/Race												
Hispanic or Latino	52	739	10	19	30	58	12	23	0	0	12	23
American Indian or Alaska Native	1	740	0	0	1	100	0	0	0	0	0	0
Asian	21	755	0	0	5	24	15	71	1	5	16	76
Black or African-American	53	738	8	15	33	62	12	23	0	0	12	23
Native Hawaiian or Other Pacific Islander	0	0	0	0	0	0	0	0	0	0	0	0
White	32	748	3	9	14	44	15	47	0	0	15	47
Two or more races	4	737	1	25	3	75	0	0	0	0	0	0
Not Indicated	0	0	0	0	0	0	0	0	0	0	0	0
Economic Disadvantage												
No	106	745	12	11	47	44	46	43	1	1	47	44
Yes	57	738	10	18	39	68	8	14	0	0	8	14
Students with Disabilities												
IEP - Yes	29	733	9	31	17	59	3	10	0	0	3	10
IEP - No	134	745	13	10	69	51	51	38	1	1	52	39
504	15	744	1	7	7	47	7	47	0	0	7	47
EL												
No	149	744	16	11	79	53	53	36	1	1	54	36
Yes	14	732	6	43	7	50	1	7	0	0	1	7

This report is NOT for public review. Distribution within your school/district must be in accordance with state and federal privacy laws, and local school board policy.

03/23/2020

SCHOOL PERFORMANCE LEVEL SUMMARY REPORT

MCAP
Maryland Comprehensive Assessment Program

CONFIDENTIAL - DO NOT DISTRIBUTE

Demonstration School 1
Demonstration District A
MARYLAND

JANUARY 2020

HIGH SCHOOL SCIENCE ASSESSMENT, 2019-2020

Purpose: This report describes group achievement in terms of average scale scores and performance levels.

	Number of Valid Scores	Average Scale Score	Performance Levels								≥ Level 4 Met or Exceeded Expectations	
			Level 2 Partially Met Expectations		Level 3 Approached Expectations		Level 4 Met Expectations		Level 5 Exceeded Expectations			
			#	%	#	%	#	%	#	%	#	%
State	**20,289**	**747**	**3,435**	**17**	**7,785**	**38**	**7,999**	**39**	**1,070**	**5**	**9,069**	**45**
District	163	743	22	13	86	53	54	33	1	1	55	34
School	6	738	2	33	3	50	1	17	0	0	1	17
Gender												
Female	3	728	2	67	1	33	0	0	0	0	0	0
Male	3	748	0	0	2	67	1	33	0	0	1	33
Ethnicity/Race												
Hispanic or Latino	0	0	0	0	0	0	0	0	0	0	0	0
American Indian or Alaska Native	0	0	0	0	0	0	0	0	0	0	0	0
Asian	0	0	0	0	0	0	0	0	0	0	0	0
Black or African-American	1	723	1	100	0	0	0	0	0	0	0	0
Native Hawaiian or Other Pacific Islander	0	0	0	0	0	0	0	0	0	0	0	0
White	5	741	1	20	3	60	1	20	0	0	1	20
Two or more races	0	0	0	0	0	0	0	0	0	0	0	0
Not Indicated	0	0	0	0	0	0	0	0	0	0	0	0
Economic Disadvantage												
No	4	739	1	25	2	50	1	25	0	0	1	25
Yes	2	736	1	50	1	50	0	0	0	0	0	0
Students with Disabilities												
IEP - Yes	1	723	1	100	0	0	0	0	0	0	0	0
IEP - No	5	741	1	20	3	60	1	20	0	0	1	20
504	0	0	0	0	0	0	0	0	0	0	0	0
EL												
No	6	738	2	33	3	50	1	17	0	0	1	17
Yes	0	0	0	0	0	0	0	0	0	0	0	0

This report is NOT for public review. Distribution within your school/district must be in accordance with state and federal privacy laws, and local school board policy.

03/23/2020

MCAP
Maryland Comprehensive Assessment Program

STUDENT ROSTER REPORT

CONFIDENTIAL - DO NOT DISTRIBUTE

Demonstration School 1
Demonstration District A
MARYLAND
JANUARY 2020

HIGH SCHOOL SCIENCE ASSESSMENT, 2019-2020

STUDENT	OVERALL SCALE SCORE	Integrated Dimensions: Physical Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts*	Integrated Dimensions: Life Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts*	Integrated Dimensions: Earth and Space Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts*
STATE	**747**	11 | 27 | 62	8 | 24 | 68	8 | 30 | 62
DISTRICT	**743**	10 | 37 | 53	6 | 28 | 66	7 | 39 | 55
SCHOOL	**738**	17 | 50 | 33	17 | 50 | 33	17 | 33 | 50
LNAME1, FNAME1	727	↔	↓	↔
LNAME10, FNAME10	748	↔	↑	↑
LNAME12, FNAME12	723	↓	↔	↓
LNAME3, FNAME3	734	↑	↔	↔
LNAME5, FNAME5	742	↔	↑	↑
LNAME8, FNAME8	753	↑	↔	↑

2 Partially Met Expectations (650-729)	3 Approached Expectations (730-749)	4 Met Expectations (750-774)	5 Exceeded Expectations (775-850)	↓ Partially Met Expectations	↔ Approached Expectations	↑ Met or Exceeded Expectations

This report is NOT for public review. Distribution within your school/district must be in accordance with state and federal privacy laws, and local school board policy.
* Numbers are percentages

03/23/2020

MCAP
Maryland Comprehensive Assessment Program

Student Name: FNAME8 LNAME8
SASID: DA00800008
Date of Birth: Demo 01/01/2020
Administration: JANUARY 2020

LSS Name: Demonstration District A
School Name: Demonstration School 1
Grade: 10

High School Science Assessment Report, 2019-2020

This report shows whether FNAME8 met high school grade band expectations in science and is on track to be college and career ready. The **Maryland Integrated Science Assessment (MISA)** is just one measure of how well your child is performing academically.

How Can You Use This Report?

Ask your child's teachers:

- What do you see as my child's strengths and areas for improvement in science?
- How can these assessment results be used to help my child make progress in science?

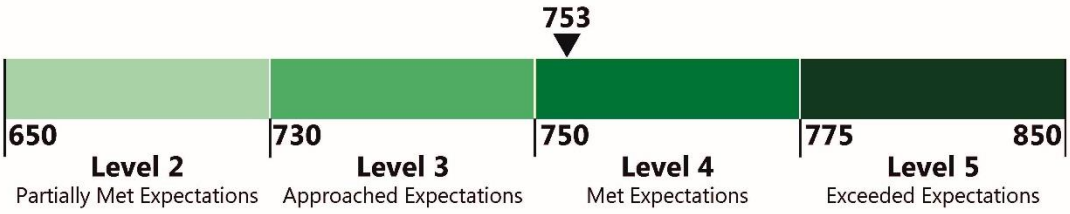
To learn more about the test and to view sample questions and practice tests, please visit: http://marylandpublicschools.org/about/Pages/DAAIT/Assessment/MISA/index.aspx.

See side 2 of this report for specific information on your child's performance in science.

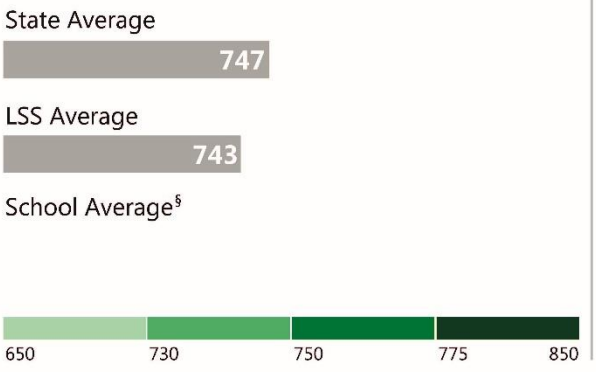
How did FNAME8 perform overall?

OVERALL STUDENT PERFORMANCE

Your student scored **753** on a scale of **650-850**, and performed at **LEVEL 4 - MET EXPECTATIONS**. Students performing at this level demonstrate a general understanding of high school appropriate Disciplinary Core Ideas, Science and Engineering Practices, and Crosscutting Concepts.



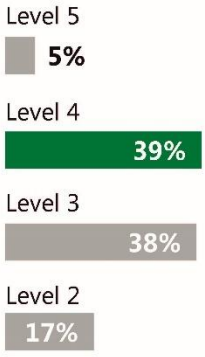
School, LSS, and State Comparisons



§ Note: Some numbers may have been left blank because fewer than ten (10) students were tested.

How Students in Maryland Performed

Percentage of students at each performance level



03/23/2020

MARYLAND STATE DEPARTMENT OF EDUCATION
EQUITY AND EXCELLENCE

Student Name: FNAME8 LNAME8

What is an Integrated Science Assessment?

The MISA is given in grades 5, 8 and high school. Each assessment integrates the **disciplinary core ideas** of physical science, life science, and earth and space science, as well as engineering, technology, and applications of science. These disciplines are then integrated with the **science and engineering practices** and **crosscutting concepts** described below:

Science and Engineering Practices: The behaviors and processes that scientists engage in to make sense of phenomena and design solutions to problems.

- Asking questions and defining problems
- Developing and using models
- Planning and carrying out investigations
- Analyzing and interpreting data
- Using mathematics and computational thinking
- Constructing explanations and designing solutions
- Engaging in argument from evidence
- Obtaining, evaluating and communicating information

Crosscutting Concepts: An organizational framework for connecting knowledge from the various discipline into a coherent and scientifically based view of the world.

- Patterns
- Cause and effect: Mechanism and explanation
- Scale, proportion, and quantity
- Systems and system models
- Energy and matter: Flows, cycles, and conservation
- Structure and function
- Stability and change
- Interdependence of science, engineering, and technology
- Influence of engineering, technology, and science on society and the natural world

The integration of these three dimensions - **disciplinary core ideas**, **science and engineering practices**, and **crosscutting concepts** - is fundamental to understanding science and central to the design of the MISA.

Further information about the performance levels at each grade band can be found online at: http://marylandpublicschools.org/about/Pages/DAAIT/Assessment/MISA/index.aspx.

How Did Your Child Perform on the Integrated Dimensions of the HS MISA?

Physical Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts

Your student performed about the same as students who **met or exceeded expectations**. Students meet expectations by demonstrating the ability to apply and integrate science and engineering practices and crosscutting concepts to the understanding of matter and its interactions, motion and stability, forces and interactions, energy and waves.

Life Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts

Your student performed about the same as students who **approached expectations**. Students meet expectations by demonstrating the ability to apply and integrate science and engineering practices and crosscutting concepts to the understanding of how the structures and processes function from molecules to organisms, the interactions, energy and dynamics of ecosystems, the inheritance and variation of traits in heredity, and the unity and diversity of biological evolution.

Earth and Space Sciences Integrated with Science and Engineering Practices and Crosscutting Concepts

Your student performed about the same as students who **met or exceeded expectations**. Students meet expectations by demonstrating the ability to apply and integrate science and engineering practices and crosscutting concepts to the understanding of Earth's place in the universe, Earth's sytems, and Earth and human activity.

LEGEND Your child performed about the same as students who: Met or Exceeded Expectations; Approached Expectations; Partially Met Expectations

How are assessment results used? Results from the assessment give your child's teacher, school, and LSS information about his/her science performance, and provide you with some insight on how your child is meeting expectations. These results never stand alone, but can be used with other assessments and class work when gauging student performance.

Learn more about Maryland's science standards

NGSS web site: https://www.nextgenscience.org/

MDK12 website: http://marylandpublicschools.org/about/Pages/DAAIT/Assessment/MISA/index.aspx

January 2020 Maryland Integrated Science Assessment (MISA)

LNAME1, FNAME1

LSS: Demonstration District A (DA)

School: Demonstration School 1 (DEM1)

Grade: 11 ID: DA00800001 DOB: 08/08/2002

MISA Scale Score: **727**

Performance Level: **Level 2 - Partially Met Expectations**

January 2020 Maryland Integrated Science Assessment (MISA)

LNAME10, FNAME10

LSS: Demonstration District A (DA)

School: Demonstration School 1 (DEM1)

Grade: 10 ID: DA00800010 DOB: 06/10/2004

MISA Scale Score: **748**

Performance Level: **Level 3 - Approached Expectations**

January 2020 Maryland Integrated Science Assessment (MISA)

LNAME12, FNAME12

LSS: Demonstration District A (DA)

School: Demonstration School 1 (DEM1)

Grade: 12 ID: DA00800012 DOB: 06/28/2002

MISA Scale Score: **723**

Performance Level: **Level 2 - Partially Met Expectations**

January 2020 Maryland Integrated Science Assessment (MISA)

LNAME3, FNAME3

LSS: Demonstration District A (DA)

School: Demonstration School 1 (DEM1)

Grade: 10 ID: DA00800003 DOB: 06/06/2003

MISA Scale Score: **734**

Performance Level: **Level 3 - Approached Expectations**

January 2020 Maryland Integrated Science Assessment (MISA)

LNAME5, FNAME5

LSS: Demonstration District A (DA)

School: Demonstration School 1 (DEM1)

Grade: 10 ID: DA00800005 DOB: 05/21/2004

MISA Scale Score: **742**

Performance Level: **Level 3 - Approached Expectations**

January 2020 Maryland Integrated Science Assessment (MISA)

LNAME8, FNAME8

LSS: Demonstration District A (DA)

School: Demonstration School 1 (DEM1)

Grade: 10 ID: DA00800008 DOB: 04/25/2004

MISA Scale Score: **753**

Performance Level: **Level 4 - Met Expectations**

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
405	394	PASS

LASTNAME1, FIRSTNAME1 A
Date of Birth: 01/01/2004
State ID: 1234567890
LEA ID: 0000123456
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
336	394	FAIL

LASTNAME2, FIRSTNAME2 B
Date of Birth: 01/01/2004
State ID: 1234567891
LEA ID: 0000123457
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
317	394	FAIL

LASTNAME3, FIRSTNAME3 C
Date of Birth: 01/01/2004
State ID: 1234567892
LEA ID: 0000123458
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
360	394	FAIL

LASTNAME4, FIRSTNAME4 D
Date of Birth: 01/01/2004
State ID: 1234567893
LEA ID: 0000123459
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
409	394	PASS

LASTNAME5, FIRSTNAME5 E
Date of Birth: 02/02/2004
State ID: 1234567894
LEA ID: 0000123460
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
349	394	FAIL

LASTNAME6, FIRSTNAME6 F
Date of Birth: 03/03/2004
State ID: 1234567895
LEA ID: 0000123461
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
335	394	FAIL

LASTNAME7, FIRSTNAME7 G
Date of Birth: 09/09/2003
State ID: 1234567896
LEA ID: 0000123462
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
387	394	FAIL

LASTNAME8, FIRSTNAME8 H
Date of Birth: 04/04/2003
State ID: 1234567897
LEA ID: 0000123463
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
395	394	PASS

LASTNAME9, FIRSTNAME9 I
Date of Birth: 02/07/2003
State ID: 1234567898
LEA ID: 0000123464
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
363	394	FAIL

LASTNAME10, FIRSTNAME10 J
Date of Birth: 11/11/2003
State ID: 1234567899
LEA ID: 0000123465
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
332	394	FAIL

LASTNAME11, FIRSTNAME11 K
Date of Birth: 12/12/2003
State ID: 1234567900
LEA ID: 0000123466
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Maryland HSA - Government

Test Date: January 2020

Student's Score	Passing Score	Pass/Fail Status
440	394	PASS

LASTNAME12, FIRSTNAME12 L
Date of Birth: 05/05/2004
State ID: 1234567901
LEA ID: 0000123467
LEA Name: Demonstration District A
LEA Number: DA
School Name: Demonstration School 1
School Number: DEM1

Appendix B. Classical Item Statistics—Operational Items

For the data in tables B-1 through B-7:

- Item Type = Type + Point Value, where Type is one of the following:
  - BCR (brief constructed-response items worth 4 points),
  - CR (constructed-response items worth 2, 3, or 4 points),
  - MSR (multi-select items worth either 1 or 2 points),
  - SR (selected-response items), or
  - TE (technology-enhanced items worth either 1 or 2 points).
- Common = whether the item appears on other forms in this administration
  - L= item is common across all forms in this administration,
  - O = item is in one or more but not all forms in this administration.
- Forms = the forms on which the item appears in this administration,
- P_Val = *p*-value,
- R_ITT = item-total correlation,
- P_BIS1 – P_BIS*n* = option-total correlations for *n* options, and
- %Omits = percentage of omitted responses.

Table B-1. Classical Item Statistics, Operational Items: HSA Government—January 2020—Forms A–C (*N* = 8,336)

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
BCR-4	O	005041	0.34	0.61					4.8
BCR-4	L	0061ES	0.09	0.54					11.4
Mean (BCR-4)			0.21	0.58					8.1
SD (BCR-4)			0.18	0.05					4.7
ECR-5	O	005STO	0.26	0.74					6.4
SR	L	005AWN	0.72	0.32	-0.22	0.32	-0.13	-0.08	0.7
SR	L	005F8Q	0.36	0.42	-0.19	0.42	-0.16	-0.10	1.0
SR	L	0053C4	0.44	0.31	-0.09	-0.26	0.31	-0.02	1.4
SR	O	005FA1	0.40	0.28	-0.15	-0.08	-0.13	0.28	1.4
SR	O	005B73	0.61	0.37	-0.11	-0.28	0.37	-0.06	1.4
SR	L	005077	0.45	0.44	-0.19	0.44	-0.22	-0.10	1.6
SR	L	0053EI	0.69	0.31	-0.19	-0.14	-0.12	0.31	1.7
SR	L	0053F4	0.57	0.43	-0.14	-0.21	-0.21	0.43	1.8
SR	L	005BAG	0.57	0.41	-0.23	-0.19	0.41	-0.13	0.7
SR	L	005B00	0.58	0.43	-0.15	-0.22	0.43	-0.20	2.6
SR	O	005B1V	0.30	0.15	0.15	-0.12	-0.06	0.06	2.7
SR	O	0065L3	0.62	0.41	-0.19	-0.19	0.41	-0.18	0.7
SR	O	0061AR	0.45	0.51	0.51	-0.23	-0.20	-0.20	0.9
SR	L	0053CV	0.44	0.47	-0.15	-0.24	-0.22	0.47	0.8
SR	L	004ZV0	0.46	0.35	-0.08	-0.23	0.35	-0.09	1.0
SR	L	0053C5	0.65	0.37	-0.16	-0.16	-0.19	0.37	0.9
SR	O	005UTR	0.42	0.24	0.01	0.24	-0.11	-0.15	0.9
SR	O	005078	0.62	0.51	-0.22	-0.27	0.51	-0.21	1.1
SR	L	0065LC	0.43	0.54	0.54	-0.25	-0.20	-0.20	1.1
SR	O	006541	0.43	0.53	-0.21	-0.28	0.53	-0.14	1.1
SR	O	005STM	0.23	0.46	-0.09	-0.13	-0.18	0.46	1.1
SR	O	005STL	0.55	0.48	-0.19	0.48	-0.29	-0.14	1.2
SR	O	005STK	0.39	0.35	-0.20	-0.14	0.35	-0.05	1.2
SR	O	005STN	0.44	0.51	-0.15	-0.14	-0.31	0.51	1.3
SR	O	005B0W	0.54	0.43	-0.25	-0.18	0.43	-0.12	1.7
SR	L	0053JF	0.25	0.38	-0.14	0.05	-0.26	0.38	1.7
SR	L	0053CI	0.37	0.30	-0.21	0.30	-0.14	0.06	1.8
SR	L	005BD7	0.19	0.19	-0.07	-0.10	0.19	0.07	1.8
SR	L	005BCI	0.31	0.35	-0.07	-0.15	-0.14	0.35	1.8

continued

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
SR	O	005BJJ	0.41	0.26	0.03	-0.15	0.26	-0.16	1.9
SR	L	005SXQ	0.29	0.36	-0.04	0.36	-0.12	-0.18	1.8
SR	L	0053D3	0.33	0.27	0.27	0.01	-0.23	-0.08	1.9
SR	L	005BF3	0.38	0.34	-0.12	-0.16	0.34	-0.08	1.9
SR	L	005FAC	0.15	0.13	0.20	0.13	-0.18	-0.10	2.0
SR	L	005BH4	0.30	0.39	-0.07	-0.17	-0.14	0.39	2.1
SR	L	005F1I	0.50	0.49	-0.23	-0.19	0.49	-0.18	2.2
SR	L	005BK8	0.64	0.47	0.47	-0.20	-0.24	-0.18	2.2
SR	O	0053DU	0.29	0.36	0.06	-0.11	0.36	-0.22	2.3
SR	L	0053AR	0.52	0.50	-0.24	-0.23	-0.18	0.50	2.3
SR	L	0065LD	0.40	0.36	-0.15	0.36	-0.13	-0.10	2.4
SR	L	005BDQ	0.55	0.46	-0.18	-0.27	0.46	-0.15	2.3
SR	L	005B24	0.40	0.42	-0.11	-0.26	0.42	-0.06	1.9
SR	O	005AUN	0.35	0.28	-0.10	0.28	-0.20	0.00	2.0
SR	L	0065KQ	0.39	0.47	-0.24	-0.14	0.47	-0.15	1.9
Mean (SR)			0.44	0.38	-0.07	-0.06	0.05	0.02	1.6
SD (SR)			0.14	0.10	0.20	0.23	0.29	0.24	0.5
TE-2	O	0063VU	0.58	0.57					0.0
TE-2	O	005Y15	0.39	0.37					0.0
TE-2	O	0060YA	0.42	0.47					0.0
TE-2	O	0089UU	0.62	0.58					0.0
TE-2	O	005UO3	0.43	0.28					0.0
Mean (TE-2)			0.49	0.45					0.0
SD (TE-2)			0.10	0.13					0.0

Table B-2. Classical Item Statistics, Operational Items: HSA Government—January 2020—Forms AA–AC (*N* = 8,347)

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
BCR-4	O	0061AS	0.26	0.63					6.3
BCR-4	L	0061ES	0.09	0.53					11.3
Mean (BCR-4)			0.17	0.58					8.8
SD (BCR-4)			0.12	0.07					3.5
ECR-5	O	005SU4	0.26	0.75					6.5
SR	L	005AWN	0.73	0.31	-0.22	0.31	-0.12	-0.07	0.6
SR	L	005F8Q	0.36	0.40	-0.18	0.40	-0.16	-0.09	0.9
SR	L	0053C4	0.44	0.29	-0.09	-0.27	0.29	0.00	1.1
SR	O	005F26	0.69	0.41	-0.16	-0.21	0.41	-0.21	1.2
SR	O	00507I	0.53	0.26	-0.12	0.26	-0.17	-0.03	1.2
SR	L	5077	0.45	0.44	-0.17	0.44	-0.23	-0.12	1.4
SR	L	0053EI	0.71	0.33	-0.20	-0.16	-0.13	0.33	1.4
SR	L	0053F4	0.57	0.44	-0.15	-0.20	-0.22	0.44	1.5
SR	L	005BAG	0.56	0.40	-0.23	-0.19	0.40	-0.12	0.6
SR	L	005B00	0.60	0.43	-0.16	-0.22	0.43	-0.19	2.3
SR	O	005SXL	0.43	0.39	0.39	-0.21	-0.18	-0.07	2.2
SR	O	007O59	0.65	0.41	-0.22	-0.18	0.41	-0.19	0.7
SR	O	0053DD	0.44	0.30	-0.01	0.30	-0.20	-0.17	0.8
SR	L	0053CV	0.45	0.47	-0.14	-0.24	-0.23	0.47	0.7
SR	L	004ZV0	0.47	0.32	-0.07	-0.22	0.32	-0.08	0.8
SR	L	0053C5	0.65	0.36	-0.16	-0.15	-0.20	0.36	0.7
SR	O	0065KZ	0.69	0.45	0.45	-0.28	-0.23	-0.15	0.8
SR	O	005EOS	0.79	0.34	-0.22	0.34	-0.15	-0.13	1.0
SR	L	0065LC	0.43	0.53	0.53	-0.25	-0.19	-0.18	1.0
SR	O	005AOP	0.44	0.47	-0.23	-0.16	-0.21	0.47	1.0
SR	O	005SU0	0.29	0.29	-0.02	-0.16	0.29	-0.11	1.1
SR	O	005SU2	0.28	0.40	-0.12	-0.17	-0.10	0.40	1.2
SR	O	005SU1	0.43	0.43	-0.21	-0.10	-0.22	0.43	1.1
SR	O	005SU3	0.42	0.51	-0.23	-0.21	0.51	-0.19	1.1
SR	O	005FDV	0.78	0.36	-0.16	0.36	-0.21	-0.13	1.8
SR	L	0053JF	0.25	0.37	-0.15	0.07	-0.27	0.37	1.9
SR	L	0053CI	0.35	0.30	-0.18	0.30	-0.15	0.05	1.9
SR	L	005BD7	0.18	0.17	-0.09	-0.08	0.17	0.09	1.9
SR	L	005BCI	0.35	0.30	-0.03	-0.16	-0.13	0.30	1.9

continued

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
SR	O	005BDO	0.30	0.42	-0.13	-0.03	-0.25	0.42	2.0
SR	L	005SXQ	0.29	0.32	-0.02	0.32	-0.11	-0.16	1.9
SR	L	0053D3	0.34	0.27	0.27	-0.01	-0.23	-0.05	2.0
SR	L	005BF3	0.41	0.32	-0.11	-0.16	0.32	-0.07	2.1
SR	L	005FAC	0.15	0.13	0.19	0.13	-0.18	-0.09	2.1
SR	L	005BH4	0.29	0.40	-0.10	-0.19	-0.11	0.40	2.3
SR	L	005F1I	0.50	0.49	-0.22	-0.19	0.49	-0.17	2.2
SR	L	005BK8	0.64	0.47	0.47	-0.20	-0.25	-0.18	2.3
SR	O	005BEU	0.37	0.30	-0.06	0.30	-0.12	-0.15	2.4
SR	L	0053AR	0.51	0.50	-0.24	-0.21	-0.19	0.50	2.5
SR	L	0065LD	0.41	0.37	-0.17	0.37	-0.14	-0.09	2.7
SR	L	005BDQ	0.56	0.46	-0.19	-0.26	0.46	-0.15	2.5
SR	L	005B24	0.39	0.41	-0.10	-0.28	0.41	-0.04	2.0
SR	O	00507E	0.40	0.46	-0.11	0.46	-0.16	-0.25	2.1
SR	L	0065KQ	0.45	0.45	-0.27	-0.12	0.45	-0.12	2.0
Mean (SR)			0.46	0.38	-0.08	-0.03	0.00	0.03	1.6
SD (SR)			0.16	0.09	0.20	0.25	0.27	0.24	0.6
TE-2	O	0063VU	0.59	0.58					0.0
TE-2	O	005Y15	0.38	0.39					0.0
TE-2	O	0060YA	0.41	0.48					0.0
TE-2	O	0089UU	0.62	0.58					0.0
TE-2	O	005Y2A	0.47	0.42					0.0
Mean (TE-2)			0.49	0.49					0.0
SD (TE-2)			0.11	0.09					0.0

Table B-3. Classical Item Statistics, Operational Items: HSA Government—January 2020—Accommodated Form X (*N* = 1,967)

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
BCR-4	O	005041	0.22	0.50					7.1
BCR-4	L	0061ES	0.03	0.35					11.2
Mean (BCR-4)			0.13	0.42					9.2
SD (BCR-4)			0.14	0.11					2.9
ECR-5	O	005STO	0.14	0.57					8.0
MSR-2	O	006UHI	0.33	0.47					0.0
MSR-2	O	006SGW	0.29	0.23					0.0
MSR-2	O	006UY6	0.27	0.38					0.0
MSR-2	O	006UG2	0.35	0.22					0.0
Mean (MSR-2)			0.31	0.32					0.0
SD (MSR-2)			0.03	0.12					0.0
SR	L	005AWN	0.67	0.22	-0.13	0.22	-0.11	-0.03	0.6
SR	L	005F8Q	0.27	0.18	-0.06	0.18	-0.06	-0.03	1.0
SR	L	0053C4	0.35	0.18	-0.08	-0.16	0.18	0.09	1.5
SR	O	005FA1	0.26	0.19	-0.10	-0.02	-0.06	0.19	1.5
SR	O	005B73	0.44	0.31	-0.04	-0.22	0.31	-0.07	1.6
SR	L	5077	0.30	0.24	-0.13	0.24	-0.11	0.05	1.6
SR	L	0053EI	0.54	0.35	-0.19	-0.08	-0.15	0.35	1.8
SR	L	0053F4	0.40	0.31	-0.11	-0.09	-0.12	0.31	1.8
SR	L	005BAG	0.41	0.25	-0.07	-0.11	0.25	-0.13	0.7
SR	L	005B00	0.46	0.27	-0.05	-0.12	0.27	-0.10	3.2
SR	O	005B1V	0.24	0.16	0.16	-0.08	0.01	-0.03	3.1
SR	O	0065L3	0.46	0.27	-0.12	-0.06	0.27	-0.12	0.8
SR	O	0061AR	0.33	0.28	0.28	-0.10	-0.08	-0.11	0.9
SR	L	0053CV	0.26	0.29	-0.16	-0.17	0.01	0.29	0.9
SR	L	004ZV0	0.36	0.18	-0.08	-0.07	0.18	-0.03	1.4
SR	L	0053C5	0.54	0.28	-0.15	-0.12	-0.08	0.28	1.0
SR	O	005UTR	0.36	0.18	-0.05	0.18	-0.08	-0.03	1.2
SR	O	5078	0.37	0.45	-0.19	-0.23	0.45	-0.06	1.3
SR	L	0065LC	0.25	0.26	0.26	-0.09	-0.07	-0.04	1.5
SR	O	6541	0.31	0.26	-0.04	-0.15	0.26	-0.02	1.4
SR	O	005STM	0.14	0.06	0.00	0.06	-0.04	0.06	1.5
SR	O	005STL	0.42	0.34	-0.13	0.34	-0.11	-0.15	1.6
SR	O	005STK	0.33	0.17	-0.08	-0.07	0.17	0.03	1.6
SR	O	005STN	0.24	0.28	-0.09	-0.10	-0.05	0.28	1.7

continued

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
SR	O	005B0W	0.40	0.26	-0.14	-0.15	0.26	0.03	1.8
SR	L	0053JF	0.14	0.18	-0.04	0.08	-0.10	0.18	1.7
SR	L	0053CI	0.29	0.13	-0.05	0.13	-0.04	0.05	1.8
SR	L	005BD7	0.21	0.09	0.02	0.00	0.09	-0.03	1.8
SR	L	005BCI	0.26	0.19	0.03	-0.11	-0.05	0.19	1.7
SR	O	005BJJ	0.33	0.15	0.05	-0.06	0.15	-0.09	1.9
SR	L	005SXQ	0.23	0.16	0.01	0.16	-0.04	-0.04	1.8
SR	L	0053D3	0.30	0.16	0.16	0.07	-0.14	-0.06	1.8
SR	L	005BF3	0.32	0.17	-0.03	-0.07	0.17	-0.02	1.9
SR	L	005FAC	0.16	0.01	0.05	0.01	0.02	-0.02	1.9
SR	L	005BH4	0.21	0.22	-0.05	-0.07	-0.01	0.22	2.0
SR	L	005F1I	0.33	0.27	-0.07	-0.12	0.27	-0.04	2.0
SR	L	005BK8	0.46	0.38	0.38	-0.13	-0.17	-0.12	2.1
SR	O	0053DU	0.26	0.09	0.00	-0.03	0.09	0.02	2.1
SR	L	0053AR	0.33	0.37	-0.17	-0.06	-0.16	0.37	2.2
SR	L	0065LD	0.28	0.22	-0.09	0.22	-0.05	-0.01	2.2
SR	L	005BDQ	0.37	0.31	-0.11	-0.18	0.31	0.00	2.3
SR	L	005B24	0.27	0.19	-0.01	-0.11	0.19	0.01	1.8
SR	O	005AUN	0.30	0.21	0.00	0.21	-0.16	-0.01	1.9
SR	L	0065KQ	0.27	0.25	0.00	-0.09	0.25	-0.13	1.8
Mean (SR)									
SD (SR)									
TE-2	O	006UFG	0.53	0.41					0.0

Table B-4. Classical Item Statistics, Operational Items: HS MISA—January 2020—Forms A, AA, AD (*N* = 6,462)

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
CR-2	L	006IGR	0.19	0.67					7.1
CR-2	O	006JVC	0.12	0.51					4.3
Mean(CR-2)			0.16	0.59					5.7
SD(CR-2)			0.05	0.11					1.9
CR-3	L	005HGP	0.15	0.48					3.6
CR-3	O	0064KQ	0.20	0.66					6.0
Mean(CR-3)			0.17	0.57					4.8
SD(CR-3)			0.04	0.12					1.7
CR-4	L	005WON	0.16	0.70					4.6
CR-4	O	006EG3	0.10	0.60					7.6
Mean(CR-4)			0.13	0.65					6.1
SD(CR-4)			0.05	0.07					2.2
MSR-1	L	006RH5	0.47	0.64					0.0
MSR-1	O	006JV9	0.19	0.40					0.0
Mean(MSR-1)			0.33	0.52					0.0
SD(MSR-1)			0.20	0.17					0.0
MSR-2	O	006JV2	0.47	0.72					0.0
MSR-2	O	0064K0	0.47	0.57					0.0
Mean(MSR-2)			0.47	0.64					0.0
SD(MSR-2)			0.00	0.10					0.0
SR	L	005K55	0.45	0.24	-0.06	-0.24	0.24	-0.04	0.5
SR	L	005H2S	0.38	0.26	-0.13	-0.23	0.26	0.05	0.5
SR	L	005H6O	0.38	0.42	-0.22	-0.09	-0.22	0.42	0.6
SR	L	005H65	0.51	0.35	-0.03	-0.27	0.35	-0.23	0.6
SR	L	006IG2	0.31	0.30	-0.04	-0.12	0.30	-0.13	0.8
SR	L	006RH9	0.59	0.36	-0.22	0.36	-0.12	-0.19	1.1
SR	O	006JV3	0.64	0.49	-0.21	-0.31	0.49	-0.17	0.7
SR	O	006JV4	0.54	0.58	0.58	-0.25	-0.32	-0.21	0.8
SR	L	005WHU	0.42	0.30	0.02	0.30	-0.31	-0.09	0.9
SR	L	005WNB	0.67	0.48	-0.18	-0.25	0.48	-0.26	0.9
SR	L	005WNE	0.57	0.61	0.61	-0.25	-0.34	-0.24	1.0
SR	L	005WO1	0.46	0.47	-0.26	-0.25	-0.14	0.47	1.0
SR	L	006R0F	0.51	0.45	0.45	-0.24	-0.25	-0.11	1.0
SR	O	006EE1	0.56	0.40	0.40	-0.12	-0.23	-0.18	1.5
SR	O	006EE6	0.27	0.23	0.23	-0.20	0.00	0.02	1.6

continued

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
SR	O	006EF8	0.67	0.50	-0.23	-0.19	-0.28	0.50	1.6
SR	O	006RGT	0.44	0.37	-0.15	0.37	-0.16	-0.13	1.7
SR	O	006EFY	0.51	0.35	-0.09	-0.12	-0.24	0.35	1.6
SR	O	006R0D	0.62	0.54	-0.19	0.54	-0.35	-0.18	1.6
SR	O	0064JR	0.59	0.61	-0.23	-0.29	-0.31	0.61	1.6
SR	O	0064JX	0.53	0.44	-0.27	0.44	-0.16	-0.16	1.7
SR	O	0064JZ	0.46	0.44	-0.15	-0.18	-0.20	0.44	1.7
Mean(SR)			0.50	0.42	-0.02	-0.07	-0.07	0.02	1.1
SD(SR)			0.11	0.11	0.28	0.27	0.28	0.29	0.4
TE-2	L	006RFO	0.44	0.33					0.0
TE-2	L	006RFS	0.43	0.51					0.0
TE-2	L	005H2Z	0.27	0.50					0.0
TE-2	O	006JUP	0.53	0.60					0.0
Mean(TE-2)			0.50	0.42	-0.02	-0.07	-0.07	0.02	1.1
SD(TE-2)			0.11	0.11	0.28	0.27	0.28	0.29	0.4

Table B-5. Classical Item Statistics, Operational Items: HS MISA—January 2020—Forms B, AB, AE (*N* = 6,480)

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
CR-2	O	006GZX	0.17	0.59					4.6
CR-2	L	006IGR	0.20	0.67					7.4
Mean(CR-2)			0.19	0.63					6.0
SD(CR-2)			0.02	0.06					2.0
CR-3	O	00570B	0.23	0.64					5.1
CR-3	L	005HGP	0.15	0.49					3.7
Mean(CR-3)			0.19	0.57					4.4
SD(CR-3)			0.06	0.11					1.0
CR-4	L	005WON	0.16	0.70					4.7
CR-4	O	0063OU	0.15	0.70					5.8
Mean(CR-4)			0.16	0.70					5.2
SD(CR-4)			0.01	0.00					0.7
MSR-1	L	006RH5	0.48	0.65					0.0
MSR-1	O	0063NZ	0.09	0.34					0.0
Mean(MSR-1)			0.28	0.49					0.0
SD(MSR-1)			0.28	0.22					0.0
MSR-2	O	006GZ9	0.35	0.57					0.0
SR	L	005K55	0.46	0.25	-0.08	-0.24	0.25	-0.03	0.5
SR	L	005H2S	0.38	0.28	-0.13	-0.23	0.28	0.02	0.6
SR	L	005H6O	0.37	0.41	-0.21	-0.09	-0.22	0.41	0.6
SR	L	005H65	0.50	0.35	-0.02	-0.28	0.35	-0.21	0.7
SR	L	006IG2	0.31	0.31	-0.05	-0.12	0.31	-0.13	0.8
SR	L	006RH9	0.59	0.36	-0.20	0.36	-0.12	-0.20	1.0
SR	O	006GXP	0.43	0.29	-0.08	0.29	-0.24	-0.04	0.8
SR	O	006GZB	0.31	0.18	0.18	-0.16	-0.06	0.05	0.9
SR	O	006GZN	0.52	0.43	0.43	-0.23	-0.16	-0.17	1.0
SR	L	005WHU	0.43	0.30	0.00	0.30	-0.29	-0.09	1.0
SR	L	005WNB	0.67	0.48	-0.19	-0.25	0.48	-0.25	1.0
SR	L	005WNE	0.57	0.58	0.58	-0.22	-0.33	-0.23	1.1
SR	L	005WO1	0.46	0.47	-0.24	-0.25	-0.15	0.47	1.1
SR	L	006R0F	0.51	0.46	0.46	-0.22	-0.25	-0.12	1.2
SR	O	0056UO	0.51	0.41	-0.12	0.41	-0.27	-0.13	1.4
SR	O	005700	0.58	0.56	0.56	-0.27	-0.29	-0.20	1.7
SR	O	0056ZV	0.44	0.34	-0.12	0.00	-0.32	0.34	1.6
SR	O	0063L6	0.31	0.36	-0.09	-0.28	-0.01	0.36	1.6

continued

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
SR	O	0063LJ	0.42	0.17	-0.18	0.17	0.09	-0.19	1.7
Mean(SR)			0.46	0.37	0.03	-0.07	-0.05	-0.02	1.1
SD(SR)			0.10	0.11	0.27	0.25	0.26	0.24	0.4
TE-1	O	0056TL	0.32	0.54					
TE-1	O	0063L7	0.38	0.56					
TE-1	O	0063LC	0.34	0.35					
Mean(TE-1)			0.35	0.48					
SD(TE-1)			0.03	0.12					
TE-2	L	006RFO	0.44	0.31					
TE-2	L	006RFS	0.43	0.50					
TE-2	L	005H2Z	0.27	0.52					
TE-2	O	006FMN	0.45	0.59					
TE-2	O	0056ZQ	0.58	0.63					
Mean(TE-2)			0.43	0.51					
SD(TE-2)			0.11	0.12					

Table B-6. Classical Item Statistics, Operational Items: HS MISA—January 2020—Forms C, AC, AF (*N* = 6,471)

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
CR-2	O	005XLL	0.09	0.53					8.6
CR-2	L	006IGR	0.20	0.68					7.3
Mean(CR-2)			0.15	0.60					8.0
SD(CR-2)			0.07	0.11					0.9
CR-3	L	005HGP	0.14	0.49					3.7
CR-3	O	006HRU	0.14	0.66					6.9
CR-3	O	006LUX	0.25	0.65					4.8
Mean(CR-3)			0.18	0.60					5.1
SD(CR-3)			0.06	0.10					1.7
CR-4	L	005WON	0.16	0.70					4.6
MSR-1	L	006RH5	0.46	0.62					0.0
MSR-1	O	006HRS	0.20	0.49					0.0
Mean(MSR-1)			0.33	0.56					0.0
SD(MSR-1)			0.19	0.09					0.0
SR	L	005K55	0.45	0.28	-0.10	-0.22	0.28	-0.06	0.4
SR	L	005H2S	0.37	0.28	-0.14	-0.20	0.28	0.02	0.5
SR	L	005H6O	0.38	0.45	-0.22	-0.12	-0.22	0.45	0.6
SR	L	005H65	0.51	0.35	-0.03	-0.26	0.35	-0.23	0.6
SR	L	006IG2	0.31	0.33	-0.04	-0.14	0.33	-0.14	0.9
SR	L	006RH9	0.60	0.36	-0.18	0.36	-0.14	-0.18	1.3
SR	O	006LUP	0.55	0.34	-0.13	-0.16	0.34	-0.16	1.0
SR	O	006LUQ	0.56	0.49	-0.15	-0.24	0.49	-0.26	1.0
SR	L	005WHU	0.41	0.34	-0.03	0.34	-0.28	-0.10	1.2
SR	L	005WNB	0.67	0.48	-0.20	-0.23	0.48	-0.24	1.2
SR	L	005WNE	0.58	0.58	0.58	-0.23	-0.30	-0.24	1.1
SR	L	005WO1	0.45	0.46	-0.24	-0.24	-0.14	0.46	1.2
SR	L	006R0F	0.49	0.47	0.47	-0.24	-0.23	-0.12	1.2
SR	O	005XGZ	0.42	0.29	-0.14	0.29	-0.17	0.01	1.6
SR	O	006HXH	0.35	0.25	-0.03	-0.18	0.25	-0.02	1.6
SR	O	006J46	0.47	0.36	-0.10	0.36	-0.22	-0.08	1.8
Mean(SR)			0.47	0.38	-0.04	-0.07	0.07	-0.06	1.1
SD(SR)			0.10	0.09	0.23	0.25	0.30	0.22	0.4
TE-1	O	006LTZ	0.19	0.43					0.0
TE-1	O	006LTT	0.57	0.45					0.0
TE-1	O	005XIF	0.22	0.11					0.0

continued

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
TE-1	O	005XH1	0.21	0.38					0.0
TE-1	O	006HQV	0.17	0.43					0.0
Mean(TE-1)									0.0
SD(TE-1)									0.0
TE-2	L	006RFO	0.44	0.34					0.0
TE-2	L	006RFS	0.42	0.52					0.0
TE-2	L	005H2Z	0.28	0.53					0.0
TE-2	O	006LUK	0.36	0.22					0.0
TE-2	O	005XJS	0.38	0.41					0.0
TE-2	O	005XJV	0.29	0.34					0.0
TE-2	O	006IYO	0.37	0.55					0.0
Mean(TE-2)									0.0
SD(TE-2)									0.0

Table B-7. Classical Item Statistics, Operational Items: HS MISA—January 2020—Accommodated Form X (*N* = 876)

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
CR-2	L	006IGR	0.03	0.52					13.5
CR-2	O	006JVC	0.02	0.41					11.2
Mean(CR-2)			0.03	0.46					12.3
SD(CR-2)			0.00	0.08					1.6
CR-3	L	005HGP	0.04	0.40					10.0
CR-3	O	0064KQ	0.07	0.53					14.5
Mean(CR-3)			0.05	0.46					12.3
SD(CR-3)			0.02	0.09					3.1
CR-4	L	005WON	0.04	0.55					12.0
CR-4	O	006EG3	0.02	0.46					15.2
Mean(CR-4)			0.03	0.51					13.6
SD(CR-4)			0.01	0.06					2.3
MSR-1	L	006RH5	0.13	0.54					0.0
MSR-1	O	006JV9	0.07	0.32					0.0
Mean(MSR-1)			0.10	0.43					0.0
SD(MSR-1)			0.04	0.15					0.0
MSR-2	O	006JV2	0.15	0.60					0.0
MSR-2	O	0064K0	0.31	0.40					0.0
Mean(MSR-2)			0.23	0.50					0.0
SD(MSR-2)			0.11	0.14					0.0
SR	L	005K55	0.40	0.16	-0.05	-0.18	0.16	0.05	0.3
SR	L	005H2S	0.31	0.07	-0.03	-0.11	0.07	0.08	0.5
SR	L	005H6O	0.23	0.17	-0.12	0.03	-0.07	0.17	0.6
SR	L	005H65	0.35	0.31	-0.03	-0.16	0.31	-0.15	0.6
SR	L	006IG2	0.25	0.09	-0.05	-0.06	0.09	0.04	1.1
SR	L	006RH9	0.44	0.28	-0.12	0.28	-0.11	-0.11	1.7
SR	O	006JV3	0.44	0.29	-0.08	-0.15	0.29	-0.11	0.8
SR	O	006JV4	0.26	0.35	0.35	-0.14	-0.11	-0.11	0.9
SR	L	005WHU	0.33	0.17	0.03	0.17	-0.20	0.01	0.8
SR	L	005WNB	0.41	0.31	-0.10	-0.13	0.31	-0.13	0.8
SR	L	005WNE	0.25	0.43	0.43	-0.12	-0.15	-0.13	0.9
SR	L	005WO1	0.25	0.33	-0.18	-0.10	-0.05	0.33	0.9
SR	L	006R0F	0.29	0.30	0.30	-0.10	-0.17	-0.02	1.0
SR	O	006EE1	0.33	0.32	0.32	-0.12	-0.12	-0.06	2.4
SR	O	006EE6	0.18	0.14	0.14	-0.02	0.01	-0.06	2.6

continued

Item Type	Anchor Status	ItemID	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
SR	O	006EF8	0.38	0.39	-0.13	-0.14	-0.14	0.39	2.7
SR	O	006RGT	0.32	0.27	-0.09	0.27	-0.04	-0.11	2.6
SR	O	006EFY	0.36	0.24	-0.05	-0.08	-0.09	0.24	2.6
SR	O	006R0D	0.37	0.39	-0.05	0.39	-0.21	-0.11	2.9
SR	O	0064JR	0.28	0.44	-0.10	-0.17	-0.12	0.44	3.4
SR	O	0064JX	0.36	0.31	-0.15	0.31	-0.06	-0.09	2.9
SR	O	0064JZ	0.24	0.20	-0.02	-0.04	-0.06	0.20	3.0
Mean(SR)			0.32	0.27	0.01	-0.02	-0.02	0.03	1.6
SD(SR)			0.07	0.10	0.18	0.18	0.16	0.18	1.0
TE-2	L	006RFO	0.35	0.15					
TE-2	L	006RFS	0.28	0.24					
TE-2	L	005H2Z	0.14	0.28					
TE-2	O	006JUP	0.32	0.43					
Mean(TE-2)			0.27	0.28					
SD(TE-2)			0.09	0.12					

Appendix C. Classical Item Statistics—Field Test Items

For the data in tables C-1 and C-2:

- Item Type = Type + Point Value, where Type is one of the following:
  - CR (constructed-response items worth 2, 3, or 4 points),
  - MSR (multi-select items worth either 1 or 2 points),
  - SR (selected-response items),
  - TE (technology-enhanced items worth either 1 or 2 points),
- P_Val = *p*-value,
- R_ITT = item-total correlation,
- P_BIS1 – P_BIS*n* = option-total correlations for *n* options, or
- %Omits = percentage of omitted responses.

Table C-1. Classical Item Statistics, Field Test Items: HSA Government—January 2020

Item Type	ItemID	N	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
CR-4	0061KB	1,845	0.29	0.70					8.3
CR-4	0064I1	1,853	0.16	0.75					7.8
CR-4	006QPC	1,824	0.23	0.67					9.5
CR-4	006QSA	1,776	0.15	0.74					12.4
CR-4	006QT9	1,778	0.27	0.73					12.4
CR-4	006QU8	1,044	0.19	0.74					17.5
Mean (CR-4)			0.22	0.72					11.3
SD (CR-4)			0.06	0.03					3.6
CR-5	005SZT	1,647	0.29	0.73					9.2
MSR-2	00851L	2,761	0.55	0.23					0.0
MSR-2	0085O7	2,790	0.62	0.55					0.0
MSR-2	0085TG	2,780	0.36	0.41					0.0
MSR-2	0085TH	2,803	0.55	0.37					0.0
MSR-2	0087XI	2,785	0.47	0.43					0.0
MSR-2	0087XW	2,764	0.58	0.50					0.0
Mean (MSR-2)			0.52	0.42					0.0
SD (MSR-2)			0.09	0.11					0.0
SR	005SZQ	8,336	0.23	0.20	-0.21	0.16	-0.21	0.20	1.8
SR	00617K	2,790	0.68	0.32	0.32	-0.15	-0.19	-0.05	3.0
SR	00617S	2,764	0.44	0.29	-0.11	-0.12	0.29	-0.15	2.0
SR	006184	2,790	0.59	0.45	0.45	-0.21	-0.24	-0.17	1.0
SR	00618Q	2,761	0.65	0.37	0.37	-0.21	-0.17	-0.14	1.0
SR	0061B5	2,780	0.43	0.48	-0.15	-0.16	-0.26	0.48	2.0
SR	0061BE	2,803	0.69	0.44	-0.20	-0.26	-0.17	0.44	2.0
SR	0061E7	2,761	0.70	0.34	0.34	-0.22	-0.09	-0.15	2.0
SR	0061EC	2,780	0.57	0.39	-0.19	0.39	-0.22	-0.12	2.0
SR	0061EJ	2,803	0.44	0.38	-0.13	0.38	-0.17	-0.11	2.0
SR	0061EX	2,764	0.43	0.45	-0.23	-0.13	-0.16	0.45	1.0
SR	0063ZM	2,761	0.63	0.49	-0.21	-0.28	-0.18	0.49	3.0
SR	006461	2,764	0.29	0.20	-0.06	-0.08	-0.05	0.20	1.0
SR	006Q8N	8,336	0.57	0.51	0.51	-0.25	-0.25	-0.16	1.6
SR	006Q8Q	8,347	0.44	0.30	-0.08	-0.14	-0.12	0.30	2.0
SR	006Q8S	8,347	0.48	0.46	0.46	-0.21	-0.17	-0.20	1.9
SR	006Q8T	8,336	0.38	0.26	0.07	-0.23	0.26	-0.13	2.0
SR	006Q8U	8,347	0.38	0.26	0.26	-0.13	-0.20	0.10	2.0
SR	006QNT	2,785	0.41	0.40	0.40	-0.18	-0.10	-0.17	2.0

continued

Item Type	ItemID	N	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
SR	006QOO	2,761	0.37	0.03	0.08	-0.10	0.03	0.03	2.0
SR	006QOY	2,785	0.31	0.25	0.02	0.25	-0.16	-0.06	1.0
SR	006QP3	2,790	0.56	0.46	-0.21	0.46	-0.20	-0.17	1.0
SR	006QQC	2,803	0.75	0.42	-0.20	-0.18	0.42	-0.20	2.0
SR	006QQK	2,790	0.31	0.38	-0.08	0.38	-0.19	-0.13	2.0
SR	006QT3	2,803	0.61	0.51	-0.16	-0.26	-0.24	0.51	3.0
SR	006QT5	2,780	0.48	0.48	-0.24	-0.17	0.48	-0.17	1.0
SR	006QTZ	2,764	0.44	0.36	-0.13	0.36	-0.17	-0.10	2.0
SR	006QU3	2,785	0.39	0.19	0.19	-0.15	-0.04	0.03	2.0
SR	006QU4	2,780	0.36	0.12	0.18	0.12	-0.26	-0.06	2.0
SR	006QU6	2,785	0.36	0.26	0.26	-0.07	-0.15	-0.05	2.0
Mean (SR)			0.48	0.35	0.04	-0.05	-0.10	0.02	1.8
SD (SR)			0.14	0.12	0.25	0.23	0.20	0.24	0.6
TE-2	006Q8V	2,761	0.54	0.37					0.0
TE-2	006QPG	2,790	0.53	0.42					0.0
TE-2	006QPJ	2,780	0.67	0.47					0.0
TE-2	006QSL	2,780	0.61	0.31					0.0
TE-2	006QSN	2,803	0.40	0.36					0.0
TE-2	006QUB	2,764	0.55	0.48					0.0
TE-2	006QUH	2,764	0.81	0.44					0.0
TE-2	006QUI	2,803	0.54	0.21					0.0
TE-2	006QUM	2,790	0.31	0.38					0.0
TE-2	006QUN	2,761	0.74	0.39					0.0
TE-2	0085D2	2,785	0.64	0.25					0.0
TE-2	0087XP	2,785	0.33	0.30					0.0
Mean (TE-2)			0.55	0.37					0.0
SD (TE-2)			0.15	0.08					0.0

Table C-2. Classical Item Statistics, Field Test Items: HS MISA—January 2020

Item Type	ItemID	N	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
CR-4	007P1D	1,878	0.29	0.67					6.0
CR-4	007P1G	1,910	0.17	0.62					5.0
Mean (CR-4)			0.23	0.65					5.5
SD (CR-4)			0.08	0.04					0.7
MSR-1	0072ZW	2,188	0.16	0.26					0.0
MSR-1	007B1Y	2,171	0.16	0.37					0.0
MSR-1	007P6K	2,202	0.14	0.10					0.0
MSR-1	007P6L	4,382	0.34	0.58					0.0
Mean (MSR-1)			0.20	0.33					0.0
SD (MSR-1)			0.09	0.20					0.0
MSR-2	007B21	2,171	0.47	0.45					0.0
MSR-2	007P68	2,202	0.30	0.26					0.0
Mean (MSR-2)			0.38	0.36					0.0
SD (MSR-2)			0.12	0.13					0.0
SR	006MA5	2,194	0.25	-0.04	0.28	-0.22	-0.04	0.01	2.0
SR	006MA6	4,390	0.41	0.19	-0.10	-0.01	0.19	-0.19	1.5
SR	006MA7	2,194	0.52	0.18	-0.15	0.18	-0.08	0.03	1.0
SR	006MA8	2,202	0.30	0.26	-0.08	-0.02	-0.16	0.26	2.0
SR	006MA9	4,390	0.37	0.39	-0.18	0.39	-0.15	-0.09	0.9
SR	006MAF	2,194	0.51	0.38	0.38	-0.28	-0.20	-0.02	1.0
SR	006MAG	4,382	0.59	0.49	-0.16	-0.25	0.49	-0.25	0.9
SR	00731H	2,202	0.51	0.55	-0.17	-0.31	-0.28	0.55	2.0
SR	007B1U	2,171	0.47	0.48	0.48	-0.30	-0.21	-0.08	2.0
SR	007B20	4,389	0.52	0.31	-0.15	0.30	-0.16	-0.07	1.8
SR	007B22	2,187	0.51	0.40	-0.11	0.40	-0.18	-0.24	1.0
SR	007B2R	4,389	0.46	0.27	-0.07	-0.23	0.27	-0.05	0.9
SR	007B2Y	2,202	0.45	0.40	-0.18	-0.24	-0.07	0.40	1.0
SR	007JL0	2,213	0.76	0.44	-0.21	-0.27	0.44	-0.16	1.0
SR	007JLH	2,190	0.38	0.23	-0.10	-0.22	0.23	0.05	1.0
SR	007JNS	4,361	0.36	0.12	0.14	0.12	-0.22	-0.07	1.5
SR	007JO0	2,171	0.45	0.38	0.38	-0.25	-0.21	-0.01	2.0
SR	007JO1	4,403	0.42	0.15	0.15	-0.19	0.14	-0.16	1.9
SR	007JO2	2,171	0.59	0.42	-0.20	0.42	-0.18	-0.20	3.0
SR	007JO4	2,213	0.62	0.50	0.50	-0.23	-0.26	-0.20	1.0
SR	007JO5	2,171	0.49	0.45	-0.20	-0.24	-0.13	0.45	2.0

continued

Item Type	ItemID	N	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
SR	007JO9	4,403	0.42	0.27	-0.09	-0.06	-0.15	0.27	1.4
SR	007P08	2,213	0.33	0.39	-0.23	-0.18	0.39	0.04	1.0
SR	007P0J	2,213	0.46	0.36	-0.02	-0.19	-0.26	0.36	1.0
SR	007P0L	2,171	0.48	0.11	-0.16	0.11	-0.20	0.15	2.0
SR	007P0N	4,403	0.53	0.27	-0.06	-0.17	0.27	-0.10	1.2
SR	007P0R	4,384	0.30	0.19	-0.14	0.19	-0.12	0.04	1.5
SR	007P0V	4,361	0.34	0.24	-0.11	0.24	-0.08	-0.05	1.5
SR	007P13	2,190	0.64	0.41	-0.20	-0.16	0.41	-0.20	2.0
SR	007P14	2,190	0.52	0.38	-0.11	0.38	-0.22	-0.14	2.0
SR	007P15	4,384	0.38	0.19	-0.10	0.09	-0.23	0.19	0.9
SR	007P4L	2,202	0.28	0.05	0.05	-0.20	0.03	0.14	0.0
SR	007P4M	4,358	0.59	0.56	-0.22	-0.25	-0.30	0.56	1.1
SR	007P4N	2,187	0.44	0.35	-0.18	-0.20	0.35	-0.07	0.0
SR	007P4P	4,373	0.59	0.37	-0.22	-0.22	0.38	-0.04	0.9
SR	007P4Q	2,171	0.34	0.18	-0.08	0.18	-0.20	0.09	2.0
SR	007P4Y	4,358	0.39	0.25	0.00	-0.18	0.25	-0.15	2.0
SR	007P4Z	2,202	0.33	0.35	0.12	-0.32	-0.27	0.35	2.0
SR	007P6A	2,202	0.35	0.49	-0.16	-0.27	-0.11	0.49	2.0
SR	007P6B	2,194	0.35	-0.11	-0.10	0.17	-0.11	0.04	2.0
SR	007P6C	4,396	0.33	0.12	-0.03	-0.06	0.12	0.00	0.7
SR	007P6D	2,188	0.55	0.56	-0.17	-0.25	-0.31	0.56	2.0
Mean (SR)			0.45	0.31	-0.05	-0.07	-0.03	0.06	1.4
SD (SR)			0.11	0.16	0.19	0.23	0.24	0.24	0.6
TE-1	006MA3	2,194	0.65	0.50					0.0
TE-1	006MA4	2,188	0.63	0.50					0.0
TE-1	006MAH	2,202	0.21	0.33					0.0
TE-1	007B2S	2,171	0.21	0.25					0.0
TE-1	007JLK	2,213	0.19	0.47					0.0
TE-1	007JNV	2,171	0.19	0.33					0.0
TE-1	007JNZ	2,190	0.16	0.29					0.0
TE-1	007OZZ	2,171	0.29	0.47					0.0
TE-1	007P0G	2,190	0.36	0.31					0.0
TE-1	007P3V	2,187	0.10	0.44					0.0
TE-1	007P6J	2,188	0.22	0.14					0.0
Mean (TE-1)			0.29	0.37					0.0
SD (TE-1)			0.18	0.12					0.0

continued

Item Type	ItemID	N	P_Val	R_ITT	P_BIS1	P_BIS2	P_BIS3	P_BIS4	%Omits
TE-2	007B1V	2,187	0.42	0.41					0.0
TE-2	007B1Z	4,373	0.57	0.52					0.0
TE-2	007B23	2,187	0.37	0.27					0.0
TE-2	007B29	2,202	0.42	0.27					0.0
TE-2	007P2N	2,171	0.50	0.48					0.0
TE-2	007P3L	2,202	0.56	0.49					0.0
TE-2	007P4R	2,187	0.17	0.32					0.0
TE-2	007P4X	2,202	0.23	0.20					0.0
TE-2	007P69	4,382	0.42	0.55					0.0
TE-2	007P6E	2,188	0.21	0.05					0.0
TE-2	007P6F	2,194	0.39	0.53					0.0
TE-2	007P6I	2,202	0.46	0.52					0.0
Mean (TE-2)			0.39	0.38					0.0
SD (TE-2)			0.13	0.16					0.0


cognia