State of Florida

Benchmarks for Excellent Student Thinking (B.E.S.T.)

2022–2023

Volume 2
Test Development

ACKNOWLEDGMENTS

This technical report was produced on behalf of the Florida Department of Education (FDOE). Requests for additional information concerning this technical report or the associated appendices should be directed to Dr. Salih Binici at the FDOE (Salih.Binici@fldoe.org).

Major contributors to this technical report include the following staff from Cambium Assessment, Inc. (CAI): Dr. Yuan Hong, Dr. Myvan Bui, Dr. Hyesuk Jang, Dr. Sherry Li, Dr. Peter Diao, Matt Gordon, Melissa Boyanton, and Zoe Dai. The major contributors from the Florida Department of Education are as follows: Dr. Salih Binici, Vince Verges, Susie Lee, Catherine Altmaier, Racquel Harrell, Sally Donelly, Shakia Johnson, Kristina Lamb, Wenyi Li, Jieling Ming, Saeyan Yun, and Yiting Yao.

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF TABLES

LIST OF FIGURES

1. INTRODUCTION

Beginning with the 2022–2023 school year, Florida's statewide, standardized assessments in English language arts (ELA) Reading, ELA Writing, Mathematics, and Mathematics end-of-course (EOC) will be aligned with the Benchmarks for Excellent Student Thinking (B.E.S.T.). The State of Florida implemented a new online assessment for operational use beginning with the 2022–2023 school year. This new assessment program, referred to as the Florida Assessment of Student Thinking (FAST), replaced the Florida Standards Assessments (FSA) in ELA Reading and Mathematics. The FAST assessments are computer adaptive, progress monitoring (PM) assessments administered three times a year. By statute, all Florida public school students are required to participate in the statewide assessments. ELA Writing 4–10 and Mathematics EOC Algebra 1 and Geometry are considered B.E.S.T. assessments and are not part of the progress monitoring FAST assessments. The FAST and B.E.S.T. were first administered to students during fall 2022, replacing the Florida Standards Assessments (FSA) in English language arts (ELA) and Mathematics. Since fall 2022, all assessments have been referred to as the FAST and B.E.S.T. Additional details on the implementation of the assessments can be found in Volume 1 of this technical report.

In spring testing windows, grades 3–8 Mathematics, grades 3–10 ELA Reading and EOC Algebra 1 and Geometry computer adaptive tests (CAT) are given to students as summative assessments. The online versions of the ELA Reading, Mathematics, Algebra 1, and Geometry assessments include the use of several technology-enhanced item types. For all online assessments, accommodated versions are available to students whose Individualized Education Plans (IEPs) or Section 504 Plans indicate such a need. ELA Writing was administered online for students in grades 4–10, with accommodations offered to students whose IEPs or Section 504 Plans stipulate the need.

The interpretation, usage, and validity of test scores rely heavily upon the process of developing the test itself. This volume provides details on the test development process of the FAST and B.E.S.T. that contributes to the validity of the test scores. Specifically, this volume provides evidence to support the following:

- The test design summary/blueprint stipulated the range of operational items from each reporting category that were required on each form. This document guided item selection and test construction for Mathematics and ELA Reading.
    - The test design summaries for both Mathematics and ELA Reading were updated during the 2022-2023 school year in order to represent the shorter test length and new reporting categories. Content Advisory Committees were conducted with educators so they could provide feedback on the overall test length, number of reporting categories, and benchmarks included within those reporting categories. The design summary now specifically states that the ELA Reading and ELA Writing components are tested and reported separately. All Mathematics and ELA tests are also administered in one session, in one day.
- The test item specifications provided detailed guidance for item writers and reviewers to ensure that the FAST and B.E.S.T. items were aligned to the standards they were intended to measure. The Test Item Specifications for both ELA and Mathematics were revised in

2021 and 2022 after the adoption of the B.E.S.T. Standards and the decision to move to a CAT. The item specifications are also updated each year as needed to document any necessary changes or clarifications that arise throughout a development cycle.

- The item development procedures employed for FAST and B.E.S.T. were consistent with industry standards.
- The development and maintenance of the FAST and B.E.S.T. item-pool plan established an item bank in which test items cover the range of measured standards, grade-level difficulties, and cognitive complexity.
- The thorough test development process contributed to the comparability of the online tests and the accommodated tests.

2. TEST SPECIFICATIONS

Following the adoption and integration of the Florida B.E.S.T. standards into the school curriculum, items and test item specifications were developed to ensure that the tests and their items were aligned to the benchmarks and grade-level expectations that they were intended to measure. FDOE and content specialists developed test-item specifications.

The FAST and B.E.S.T. test-item specifications are based on the Florida B.E.S.T. Standards and the Florida course descriptions. The specifications are a resource that defines the content and format for the test and test items for item writers and reviewers. Each grade-level and course specifications document indicates the alignment of items with the Florida B.E.S.T. Standards and also serves to provide all stakeholders with information about the scope and function of the FAST and B.E.S.T. In addition to these general guidelines, specifications for FAST ELA Reading and ELA Writing also include guidelines for developing reading and writing passages and prompts, such as length, type, and complexity.

2.1 BLUEPRINT DEVELOPMENT PROCESS

A test design summary/blueprint for each assessment specifies the number of items, item types, and reporting categories.

The blueprint construction for the FAST and B.E.S.T. in ELA and Mathematics is evidenced by the ELA and Mathematics Test Design Summary documents found at https://fsassessments.org/. These documents were created using Florida's course descriptions as the basis for the design. The course descriptions can be found on the CPALMS website at http://www.cpalms.org/Public/search/Course.

After the decision was made to switch ELA Reading and Mathematics to a computer adaptive test (CAT), Content Advisory Committee (CAC) meetings were held with educators to propose and approve the revised blueprints. An in-person CAC meeting was held to discuss the blueprints for ELA Reading and Mathematics grades 3-8 in April 2022. There was virtual CAC held for Mathematics and EOC in September 2022. In December 2022, a virtual meeting was held for both subjects to discuss potentially shortening the blueprints (although the decision was to not do so). The blueprints were also discussed at the Technical Advisory Committee (TAC) meetings in Summer and November 2022.

The reporting categories for ELA Reading were derived from the applicable "cluster" naming convention in the Florida B.E.S.T. Standards, and the percentages of the reporting categories within the tests were derived from the number, complexity, and breadth of the standards to be assessed. Vocabulary standards were folded in with the Reading Across Genres Standards to create the Reading Cross Genres & Vocabulary reporting category. Guidelines for the weight of each reporting category for FAST ELA Reading were determined by Florida's Technical Advisory Committee (TAC). The TAC advised FDOE that to avoid "statistical noise" generated from the items scored in a small reporting category, a minimum of 15% of the total raw score points should be derived from each reporting category.

The reporting categories for Mathematics were also derived from the "domain" naming convention in the Florida B.E.S.T. Standards. As with ELA Reading, if a Mathematics domain has too few

standards, two or more domains might be combined to make the reporting category 15% of the raw score points of that grade's assessment.

The benchmark information provides benchmark clarification statements, assessment limits, stimulus attributes, response attributes, prior knowledge, and a sample item for each benchmark that could be assessed.

Detailed descriptions for the constructs of the reporting categories are presented in Appendix A for ELA Reading and Appendix B for Mathematics and the EOCs.

2.1.1 Target Blueprints

Test blueprints provided the following guidelines:

- Length of the test (duration and number of items)
- Content areas to be covered and the acceptable range of items within each content area or reporting category
- Acceptable range of item difficulty for the specified grade level
- Approximate number of field-test items, if applicable
- Descriptions of test item types

This section provides only a summary of the blueprints. Detailed blueprints for each content level are presented in Appendix E for ELA, and Appendix F for Mathematics and the Mathematics EOCs.

In all grades and subjects, the assessments are administered as Computer Adaptive Tests (CAT). grades 3–10 ELA Reading, grades 3–8 Mathematics, and the Mathematics EOC assessments (Algebra 1 and Geometry) are administered online. Additionally, ELA Writing is administered online for grades 4–10. In spring 2023, typed written response accommodations were provided for students taking ELA Writing assessments in grades 4–10; therefore, responses from these students were collected online. For grades and subjects testing online, accommodations are provided if indicated by a student's IEP or Section 504 Plan.

Table 1 displays the blueprint for total test length by grade and subject or course. Each year, approximately 6–10 items on all tests are field-test items and are not used to calculate a student's score. Table 2 displays the number of operational and field-test items available in the spring 2023 item pool. ELA Writing items are not included in the item counts listed for ELA Reading tests.

Table 1: Blueprint Test Length by Grade and Subject or Course

Subject/Course	Grade	Total Number of Items
ELA Reading	3	36–40
	4	36–40
	5	36–40
	6	36–40
	7	36–40

	8	36–40
	9	36–40
	10	36–40
Mathematics	3	35
	4	35
	5	35
	6	36
	7	36
	8	36
Algebra 1		45
Geometry		45

Table 2: Number of Items Available in the Spring 2023 Item Pool by Grade and Subject or Course

Subject/Course	Grade	Number of Operational Items	Total Item Counts in the Field-Test Pool	Total Items Counts in the Field Test and Operational Pool
ELA Reading	3	158	273	431
	4	133	392	525
	5	151	375	526
	6	139	298	437
	7	173	292	465
	8	167	242	409
	9	165	325	490
	10	204	241	445
Mathematics	3	141	367	508
	4	132	271	403
	5	125	406	531
	6	182	349	531
	7	214	267	481
	8	212	139	351
Algebra 1		234	99	333
Geometry		199	177	376

Reporting categories were used to more narrowly define the topics assessed within each content area. Individual scores on reporting categories provide information to help identify areas in which a student may have had difficulty. **Error! Not a valid bookmark self-reference.** and Table 4 provide the percentage of operational items required in the blueprints by content strands, or reporting categories, for each grade level or course. The percentages shown represent an acceptable

range of item counts. As many of these items in the ELA Reading component were associated with passages, flexibility was necessary for test construction for practical reasons. The ELA Writing component prompt was not included in these blueprints. Table 5 and Table 7 provide the reporting categories for Mathematics grades 3-8 and Mathematics EOC, while Table 6 provides the percentage of operational items required in the blueprints by content strands, or reporting categories, for Mathematics EOC.

Table 3: Blueprint Percentage of Test Items Assessing Each Reporting Category in ELA Reading

Grade	Reading Prose and Poetry	Reading Informational Text	Reading Across Genres & Vocabulary
3	25–35%	25–35%	35–50%
4	25–35%	25–35%	35–50%
5	25–35%	25–35%	35–50%
6	25–35%	25–35%	35–50%
7	25–35%	25–35%	35–50%
8	25–35%	25–35%	35–50%
9	25–35%	25–35%	35–50%
10	25–35%	25–35%	35–50%

Table 4: Blueprint Percentage of Test Items Assessing Each Reporting Category in Mathematics

Grade	1*	2*	3*	4*
3	23–29%	23–29%	23–29%	23–29%
4	31–37%	31–37%	31–37%	
5	23–29%	23–29%	23–29%	23–29%
6	33–42%	25–36%	25–36%	
7	25–31%	22–31%	22–28%	22–28%
8	22–28%	22–28%	25–31%	22–28%

*See Table 5 for the reporting category names.

Table 5: Reporting Categories Used in Mathematics

Grade	Reporting Category
3	Number Sense and Additive Reasoning Number Sense and Multiplicative Reasoning Fractional Reasoning Geometric Reasoning, Measurement, and Data Analysis and Probability
4	Number Sense and Operations with Whole Numbers Number Sense and Operations with Fractions and Decimals Geometric Reasoning, Measurement, and Data Analysis and Probability
5	Number Sense and Operations with Whole Numbers Number Sense and Operations with Fractions and Decimals

	Algebraic Reasoning Geometric Reasoning, Measurement, and Data Analysis and Probability
6	Number Sense and Operations Algebraic Reasoning Geometric Reasoning, Data Analysis, and Probability
7	Number Sense and Operations and Algebraic Reasoning Proportional Reasoning and Relationships Geometric Reasoning Data Analysis and Probability
8	Number Sense and Operations and Probability Algebraic Reasoning Linear Relationships, Data Analysis, and Functions Geometric Reasoning

Table 6: Blueprint Percentage of Test Items Assessing Each Reporting Category in Mathematics EOC

Course	1*	2*	3*
Algebra 1	31–38%	31–38%	31–38%
Geometry	33–40%	27–33%	33–40%

*See Table 7 for reporting category names.

Table 7: Reporting Categories Used in EOC

Course	Reporting Category
Algebra 1	Expressions, Functions and Data Analysis Linear Relationships Non-Linear Relationships
Geometry	Logic, Relationships, and Theorems Congruence, Similarity, and Constructions Measurement and Coordinate Geometry

The FAST ELA Reading blueprint also included specifications for the genres of text presented in the passages. Two main types of text were used: literary and informational. Table 8 provides target percentages of the test passages assessing each type of text.

Table 8: Blueprint Percentage of Reading Passage Types by Grade

Grades	Informational	Literary
3–5	50%	50%
6–8	50%	50%
9–10	50%	50%

2.2 CONTENT-LEVEL AND PSYCHOMETRIC CONSIDERATIONS

In addition to the test blueprints, several content-level and psychometric considerations were used in the development of the FAST and B.E.S.T. Content-level considerations included the following:

- Selected items addressed a variety of topics.
- Identified correct answer or key was correct.
- Each item had only one correct response (some technology-enhanced items did, in fact, have more than one correct answer, and these items were reviewed to confirm that the number of correct answers matched the number asked for in the item itself).
- Identified item content or reporting category was correct.
- Items were free from typographical, spelling, punctuation, or grammatical errors.
- Items were free of any bias concerns and did not include topics that stakeholders might find offensive.
- Items fulfilled style specifications (e.g., italics, boldface, etc.).
- Items marked as do-not-use (DNU) were not selected.

Psychometric considerations included the following:

- A reasonable range of item difficulties was included.
- *p*-values for MC and CR items were reasonable and within specified bounds.
- Corrected point-biserial correlations were reasonable and within specified bounds.
- No items with negative corrected point-biserial correlations were used.
- Item response theory (IRT) *a*-parameters for all items were reasonable and greater than 0.50.
- IRT *b*-parameters for all items were reasonable and between –2 and 3.
- For MC items, IRT *c*-parameters were less than 0.40.
- Few items with model fit flags were used.
- Few items with differential item functioning (DIF) flags were used.

More information about *p*-values, corrected point-biserial correlations, IRT parameters, and DIF calculations can be found in Volume 1 of this report. The spring 2023 FAST and B.E.S.T. tests were calibrated and equated to the IRT-calibrated item pool. More details about calibration, equating, and scoring can be found in Volume 1 of this technical report.

3. ITEM DEVELOPMENT PROCEDURES

The item development procedures employed by CAI for the FAST and B.E.S.T. tests were consistent with industry practice. Just as the development of Florida's content and performance standards was an open, consensus-driven process, the development of test items and stimuli to measure those constructs was grounded in a similar philosophy.

Item development began with the following guidelines: the FAST and B.E.S.T. item specifications; the Florida B.E.S.T. Standards; language accessibility, bias, and sensitivity guidelines; editorial style guidelines; and the principles of universal design. These guidelines ensured that each aspect of a Florida item was relevant to the measured construct and was unlikely to distract or confuse test takers. In addition, these guidelines helped ensure that the wording, required background knowledge, and other aspects of the item were familiar across identifiable groups.

The principles of universal design of assessments mandate that tests are designed to minimize the impact of construct-irrelevant factors in the assessment of student achievement, removing barriers to access for the widest range of students possible. The following seven principles of universal design, as clearly defined by Thompson, Johnstone, & Thurlow (2002), were applied to the FAST and B.E.S.T. development:

1. Inclusive assessment population
2. Precisely defined constructs
3. Accessible, non-biased items
4. Amenable to accommodations
5. Simple, clear, and intuitive instructions and procedures
6. Maximum readability and comprehensibility
7. Maximum legibility

CAI applied these universal design principles in the development of all test materials, including tasks, items, and manipulatives. Test development specialists receive extensive training in item development. At every step of the review process, adherence to the principles of universal design was confirmed.

The application of universal design (UD) principles as defined by Thompson, Johnstone, & Thurlow (2002) helps develop assessments that are usable to the greatest number of test takers, including Students with Disabilities (SWDs) and English language learners (ELLs).

As documented in this technical report, the item development procedures implemented for the Florida tests are consistent with industry practice. Specifically, Florida implements the UD principles throughout every stage of the assessment development process (i.e., initial design, item development, field testing, and implementation) to minimize the need for individual accommodations. As noted by Shaftel et al. (2015), under UD principles, accessibility is integral to the item development processes, thus minimizing access barriers associated with the tests themselves to the greatest extent possible for all students, including SWDs and ELLs.

Test development specialists receive extensive training in item development, including instruction on the UD principles and guidance on designing accessible content. Adherence to the UD principles is confirmed at every step of the review process so that the test maximizes readability, legibility, and compatibility with accommodations. Checklists that align to the Council of Chief State School Officers (CCSSO) Principles for High Quality Summative Assessment are used at each phase of the development cycle. As described in the Statewide Assessment Program Information Guide (FDOE, 2019), the processes of item development and test construction are carefully guided and include many quality-control (QC) measures such as:

- Item content on accommodated forms matches item content as administered online to the extent possible (e.g., wording, graphics, paragraph breaks, option order) via multiple rounds of content reviews. Note that there are some interactions that will have accommodated form-specific language, such as equation and table match items. This additional language is needed to guide students on how to appropriately answer some items on accommodated forms.
- The student sees two-page items on an even then odd-numbered page simultaneously, just as they would see the entire item on one screen. Appropriate language is used for directives on the accommodated forms.

In terms of software that supports the item development process, CAI's Item Tracking System (ITS) served as the technology platform to efficiently carry out any item and test development process. ITS facilitated the creation of the item banks, item writing and revision, cataloging of changes and comments, and export of documents (items and passages). ITS enforced a structured review process, ensuring that every item that was written or imported underwent the appropriate sequence of reviews and signoffs; ITS archived every version of each item along with reviewer comments throughout the process. ITS also provided sophisticated pool management features that increased item quality by providing real-time, detailed item inventories and item use histories. Because ITS had the capabilities to be configured to import items in multiple formats (e.g., Microsoft Word, Excel, XML), CAI was able to import items from multiple sources. To support online test delivery, ITS had a unique Web Preview feature that displayed items exactly as they were also presented to students, using the same program code used in CAI's Test Delivery System (TDS). An online test does not have a blueline (print approval) process like a PBT, and this feature provided an item-by-item blueline capability.

Prior to test administration, a series of user acceptance testing is performed on all approved platforms to ensure that items are rendered as expected and have similar appearance across platforms to minimize potential device effects.

Rigorous review is in place to ensure that the content of the item on accommodated forms matches the content of the item as administered online (e.g., wording, graphics, paragraph breaks, option order).

The next section describes the item sources for FAST and B.E.S.T. and the subsequent sections outline the procedure used for the development and review of new items and the alignment of existing items.

3.1 SUMMARY OF ITEM SOURCES

Items for the spring 2023 FAST and B.E.S.T. came from multiple sources as outlined here.

New Items Written by CAI

New field-test items were included in the spring 2023 item pool, and these items will be used on future FAST and B.E.S.T. test forms. The newly developed field-test items were written for the Florida-specific item bank (denoted as FAST and B.E.S.T. item bank items). Mathematics and ELA items were written by CAI content experts or by trained partners. All items undergo a rigorous process of preliminary, editorial, and senior review by CAI and FDOE's Test Development Center (TDC) content teams, who followed appropriate alignment, content, and style specifications. All of these items were also reviewed by panels of Florida educators and citizens for content accuracy, and to ensure that the test items were fair, unbiased, and included topics acceptable to the Florida public.

3.2 ITEM TYPES

One of the important features of the online FAST and B.E.S.T. is the administration of technology-enhanced items. Generally referred to as Machine-Scored Constructed Response (MSCR) items, these include a wide range of item types. MSCR items require students to interact with the test content to select, construct, and/or support their answers.

Table 9 and Table 10 list the ELA Reading, Mathematics, and EOC item types, and provide a brief description of each. For accommodated forms, some of these items must be modified or replaced with other items that assess the same standard and can be scanned and scored electronically. Please see the test design summary/blueprint documents or the test item specifications for specific details. Additional information about the item types can be found in Appendix C for ELA Reading, and Appendix D for Mathematics and Mathematics EOC. Examples of various item types can be found in Appendix E.

Table 9: ELA Reading Item Types and Descriptions

Response Type	Description
multiplechoice (MC)	Student selects one correct answer from a number of options.
multipleselect (MS)	Student selects all correct answers from a number of options.
tablematch (MI)	Student checks a box to indicate if information from a column header matches information from a row. On accommodated forms, the student fills in a bubble to indicate if information from a column header matches information from a row.
hottext (HT)	Student is directed to either select or use the drag-and-drop feature to use text to support an analysis or make an inference. On accommodated forms, the student fills in bubbles to indicate which sentences are correct.
multiplechoice, hottextselectable (Two-part HT)	Student selects the correct answers from Part A and Part B. Part A is a multiple-choice or a multiselect item, and Part B is a selectable HT item.

Response Type	Description
evidence-based selectedresponse (EBSR)	Student selects the correct answers from Part A and Part B. Part A often asks the student to make an analysis or inference, and Part B requires the student to use text to support Part A.

Table 10: Mathematics and Mathematics EOC Item Types and Descriptions

Response Type	Description
multiplechoice (MC)	Student selects one correct answer from four options.
multipleselect (MS)	Student selects all correct answers from a number of options.
edittaskinlinechoice (ETIC)	Student identifies an incorrect word, phrase, or blank and chooses the replacement from a number of options. This includes items with one or more ETIC interactions. On accommodated forms, the student fills in a bubble to indicate the correct number, word, or phrase that should replace a blank or a highlighted number, word, or phrase.
grid (GI)	Student selects numbers, words, phrases, or images and uses the drag-and-drop feature to place them into a graphic. This item type may also require the student to use the point, line, or arrow tools to create a response on a graph.
hottext (HT)	Student is directed to either select or use the drag-and-drop feature to use text to support an analysis or make an inference. On accommodated forms, the student fills in bubbles to indicate which sentences are correct.
equation (EQ)	Student uses a keypad with a variety of mathematical symbols to create a response. On accommodated forms, the student fills in bubbles indicating numbers and mathematical symbols to create a response. Students respond in response grids in which they write their answer in the boxes at the top of the grid, then fill in the corresponding bubble underneath each box.
tablematch (MI)	Student checks a box to indicate if information from a column header matches information from a row. On accommodated forms, the student is directed to fill in a bubble that matches a correct option from a column with a correct option from a row.
multi-interaction (MULTI)	This is an item that contains more than one response type. It could contain more than one of the same interaction type (except for multiple combinations of ETIC), or a combination of interaction types.

3.3 COGNITIVE LABORATORIES

In a recent United States Department of Education (ED)-funded grant report investigating the accessibility of computerized assessments, Shaftel et al. (2015) point out that technology-enhanced (TE) items present greater accessibility barriers than traditional item types on accommodated tests, and that they should be examined to ensure that no construct-irrelevant variance is introduced. If some aspect of the technology impedes, or advantages, students in their responses to items, this could affect item responses and inferences regarding abilities on the measured construct.

Florida assessments are delivered by the same test delivery system as Smarter Balanced Assessment Consortium (SBAC), therefore, research evidence on the SBAC platform can also be generalizable to Florida assessments. Two types of research were conducted for SBAC: (1) usability studies on system tools and features; and (2) cognitive lab studies evaluating validity of various item types. Findings show that (1) various aspects of the test delivery system (e.g., tools, navigation, directions) provide students equitable access to the assessed content; and (2) TE item types do not introduce construct-irrelevant variance into scores. The full research report is

provided in Volume 7 of the *Florida Standards Assessments 2014–2015 Technical Report*, which was included in an earlier submission for peer review. In addition, cognitive labs on Florida items will be conducted in winter 2023. The goal of the study is to evaluate whether Florida items provide valid measures of students' mastery of the intended constructs.

3.4 ITEM TRANSLATIONS TO BRAILLE FORMAT

As is noted in Allman (2009), it is common that portions of a test may need to be modified in order to be translatable to braille format. Modifications may include substituting word, reformatting the layout of the item, and replacing untranslatable items with others of equal weight, content, and difficulty. As Winter (2010) acknowledges, this can pose a challenge to comparability, but this accommodation is needed for students with disabilities to properly demonstrate the knowledge, skills, and abilities the construct represents.

Florida uses a rigorous process, outlined in the FAST and B.E.S.T. *Production Specifications*, when creating the braille translations of its summative tests and works with the Florida Instructional Materials Center for the Visually Impaired (FIMC-VI) and the American Printing House for the Blind (APH), both of which are leaders in the industry. Both FIMC-VI and APH follow practices determined by the Braille Authority of North America (BANA).

When forms are translated into braille, our contractors ensure that the braille forms match the regular print forms and make exceptions only when modifications for the braille reader are necessary. For instance, sometimes the item directions need to be modified for the braille reader instructing them to write in the letter instead of filling in the bubble. We also provide both UEB-Nemeth and Unified English Braille (UEB)-Technical versions of Mathematics tests, and for all tests we provide both contracted and uncontracted versions to ensure that visually impaired students have the type of braille they read available. This means that in some cases, four braille transcriptions are made for each grade and subject: *UEB-Nemeth Uncontracted*, *UEB-Nemeth Contracted*, *UEB-Technical Uncontracted*, *UEB-Technical Contracted*. We ensure that the students who read braille are tested and challenged at the same level as their sighted peers. By working with FIMC-VI and APH, we ensure that all tests are reviewed and proofread by certified braille transcribers/proofreaders and teachers of the visually impaired that have vast experience and knowledge regarding students in this demographic. If modifications are made, a subject content specialist must approve any suggestions made by FIMC-VI and APH. Our content team ensures that the information vital to the item is retained in the braille format and that the student who reads braille is not given either an advantage or disadvantage.

When transcribing pictures, cartoons, and graphics, images are either described or made in a tactual format for the braille reader, or with permission from content specialists, are sometimes omitted from the test if they do not provide any additional information. If graphics are described, we often use the descriptions already created for text-to-speech, which all students have access to. If tactile graphics are created, they are kept as true to the original as possible. When deviation is needed, we comply with best practices in the field. Examples are as follows:

- Extraneous details such as decorative pictures, icons, or sections of a map that are not needed for the item are sometimes omitted—as the amount of information that can be interpreted through fingers is less than the amount of information the eye can process.

- Occasionally, especially with three-dimensional figures represented as two-dimensional drawings, graphics are too complex to be created tactually and description alone either would not provide enough information or would give away the answer. In situations such as this, we develop manipulatives of the three-dimensional figures with specific directions to the Test Administrator on how to present them.

3.5 DEVELOPMENT AND REVIEW PROCESS FOR NEW ITEMS

3.5.1 Development of New Items

CAI developed field-test items to be embedded in the FAST and B.E.S.T. operational tests. As part of the standard test development process, item writers followed the guidelines in FDOE's approved Test Item Specifications and the Test Design Summary/Blueprint.

CAI staff used the Test Item Specifications to train qualified item writers, each of whom had prior item-writing experience. The item writers were trained at CAI item-writing workshops or had previous training on writing MC and CR items. CAI content area assessment specialists worked with TDC content leads to review measurement practices in item writing and interpret the meaning of the Florida Standards and benchmarks as illustrated by the Test Item Specifications documents. This information, along with the purpose of the assessment, was explained to the item writers. Sample item stems that are included in the specifications documents served as models for the writers to use in creating items to match the Standards. To ensure that the items tapped the range of difficulty and taxonomic levels required, item writers use a method based on Webb's cognitive demands (Webb, 2002) and DOK levels.

Item writing and passage selection were guided by the following principles for each of the item types. When writing items, item writers were trained to develop items that

- have an appropriate number of correct response options or combinations;
- contain plausible distractors that represent feasible misunderstandings of the content;
- represent the range of cognitive complexities and include challenging items for students performing at all levels;
- are appropriate for students in the assigned grade in terms of reading level, vocabulary, interest, and experience;
- are embedded in a real-world context, where indicated;
- do not provide answers or hints to other items in the set or test;
- are in the form of questions or directions for task completion;
- use clear language and avoid negative constructions unless doing so provides substantial advantages; and
- are free of ethnic, gender, political, socioeconomic, and religious bias.

Similarly, reading passages should

- represent literary (fiction), informational (nonfiction), and practical selections (e.g., nontraditional pieces, including tables, charts, glossaries, indexes);

- provide students with the opportunity to interact with complex, authentic texts that may employ a variety of different structures;
- include multimedia elements when appropriate;
- be of high interest and appropriate readability for the grade level;
- be of appropriate length for the grade level;
- include topics that are in alignment with sensitivity guidelines;
- be free of ethnic, gender, political, and religious bias;
- not provide answers or hints to other items in the test; and
- include real-world texts (e.g., consumer or workplace documents, public documents such as letters to the editor, newspaper and magazine articles, thesaurus entries) to the extent possible.

When selecting passages, word count, readability, and text complexity are used in conjunction with other aspects of the passages (level of interest, accessibility of the topic, thematic elements) to determine appropriateness for a particular grade level. Table 11 provides the guidelines used in FAST ELA Reading.

Table 11: Word Counts and Readabilities of Reading Passages in FAST ELA Reading

Grade	Word Count (approximate)	Lexile Range (approximate)
3	100–700	450–900
4	100–900	770–1050
5	200–1000	770–1050
6	200–1100	955–1200
7	300–1100	955–1200
8	350–1200	955–1200
9	350–1300	1080–1400
10	350–1350	1080–1400

In FAST ELA Reading, the texts are categorized as either informational or literary. *Informational texts* inform the reader and include the following types of publications:

- Exposition: informational trade books, news articles, historical documents, essays
- Persuasive text: speeches, essays, letters to the editor, informational trade books
- Procedural texts and documents: directions, recipes, manuals, contracts

Literary texts enable the reader to explore other people's experiences or to simply read for pleasure and include the following genres:

- Narrative fiction: historical and contemporary fiction, science fiction, folktales, legends, and myths and fables
- Literary nonfiction: personal essays, biographies/autobiographies, memoirs, and speeches
- Poetry: lyrical, narrative, and epic works; sonnets, odes, and ballads

Department Item Review and Approval

After internal review, the sets of items were reviewed by content specialists at the TDC. If needed, CAI and TDC content staff discussed requested revisions, ensuring that all items appropriately measured the Florida Standards. The items were then revised by CAI and brought to Florida bias, sensitivity, and content committees for review. After any final adjustments were made to the items, including an editorial review conducted by the TDC, the TDC provided a decision for each item: *Accept as Appears*, *Accept as Revised*, or *Reject*. Items that were approved by the TDC were subsequently web-approved and placed on field-test forms.

Committee Review of New Items

All items generated for use on the FAST and B.E.S.T. were required to pass a series of rigorous reviews before they could appear as field-test items on operational test forms. The items were reviewed by three committees—the Bias Committee, the Community Sensitivity Committee, and the Content Item Review Committee.

The Bias and Sensitivity Committees reviewed items for potential bias and controversial content. These committees consisted of Florida reviewers who were selected to ensure geographic and ethnic diversity. These committees ensure that items

- present racial, ethnic, and cultural groups in a positive light;
- do not contain controversial, offensive, or potentially upsetting content;
- avoid content familiar only to specific groups of students because of race or ethnicity, class, or geographic location;
- aid in the elimination of stereotypes; and
- avoid words or phrases that have multiple meanings.

The TDC and CAI reviewed the Bias and Sensitivity Committees' feedback and conveyed any issues to the attention of the Content Item Review Committee.

The Content Item Review Committee consisted of Florida classroom teachers or content specialists by grade for each subject area. The primary responsibility of the committee members was to review all new items to ensure that they were free from such flaws as (a) inappropriate readability level, (b) ambiguity, (c) incorrect or multiple answer keys (although some item types may include multiple answer keys by design), (d) unclear instructions, and (e) factual inaccuracy. These items were approved, approved with modifications, or rejected. Only approved items were added to the item pool for the field-test stage.

3.5.2 Rubric Validation

After items were field-tested, the rubric used for scoring MSCR items was validated by a team of grade-level Florida educators. These individuals reviewed the machine-assigned scores for CR items based on the scoring rubrics and either approved the scoring rubric as it appeared on the field test or suggested revisions to the scoring based on their interpretation of the item task and the rubric. The rubric validation meeting occurred in May 2023 in-person in Tallahassee, Florida.

Similar to the items field-tested in previous years, rubrics were reviewed in one of two ways: items with simpler rubrics were reviewed via frequency tables of all student responses, while items with more complex rubrics were reviewed in 45-response samples.

Items with complex rubrics include grid (GI) items, hot text (HT) draggable items, equation (EQ) items with full keypads, text entry natural language (NL) items, and Multi-Interaction (MULTI) items containing at least one of the preceding response types.

Items with simple rubrics include edit task choice and edit task in line choice (ETIC) items, hot text (HT) selectable items, matching (MI) items, equation (EQ) items with simple numeric keypads, multiple choice and hot texts electable (Two-part HT) items, and any Multi-Interaction (MULTI) items comprised entirely of the preceding response types.

Multiple choice (MC) items, multiple select (MS) items, and Evidence-Based Selected Response (EBSR) items do not go through rubric validation.

Prior to the rubric validation meeting, CAI staff selected a sample of 45 student responses for each item with complex rubrics. The sample consisted of the following data:

- 15 responses from students who performed as expected on the item given their overall performance
- 15 responses from students who were predicted to perform well on the item given their overall performance, but instead performed poorly on the item
- 15 responses from students who were predicted to perform poorly on the item given their overall performance, but instead performed well on the item

For items with simple rubrics, CAI staff generated frequency tables that contained all student responses for each item. Frequency tables were generated out of CAI's Database of Record (DOR).

The Rubric Validation Committee reviewed 45 responses for every item with a complex rubric, having the option to approve the score or suggest a different score based on the committee's understanding of the rubric. For items with simple rubrics, the committee members were shown each item, along with the correct response and the most frequently selected incorrect responses. TDC and CAI staff ensured that the committee was scoring consistently. The committee meetings used the following procedures:

- All committee members were given a laptop allowing them to respond to the items the way a student would be able to respond in a live test.
- Each item was displayed with a projector.

- The committee discussed how to answer the item and how each point was earned.
- For items with complex rubrics, each of the 45 student responses and machine-assigned scores were displayed with a projector.
- For items with simple rubrics, the item was displayed with a projector, along with the correct response and the most frequently selected incorrect responses.
- If the committee members reached a consensus that a score was incorrect, the committee proposed modifications to the rubric.
- CAI rescored the responses using the revised rubric.
- CAI reviewed the responses that received changed scores to determine if they were correctly scored.
- The TDC reviewed the rescored responses and approved the rubric.

If any scores changed based on the Rubric Validation Committee review, CAI staff revised the machine rubric and rescored the item. After the item was rescored, CAI staff reviewed at least 10% of responses for which the score changed. This review ensured that committee suggestions were honored, that the item was scored consistently, and that no unintended changes in scoring occurred because of the revision to the machine rubric. CAI staff reviewed changes with TDC staff, and TDC staff had one final opportunity to revise the rubric or approve or reject the item.

The approved items were embedded into the spring operational test forms. At the end of the testing window, CAI conducted classical item analysis on these field-test items to ensure that the items functioned as intended with respect to the underlying scales. CAI's analysis program computed the required item and test statistics for each MC and CR item to check the integrity of the item and to verify the appropriateness of the difficulty level of the item. Key statistical analyses included item discrimination, distractor analysis, item difficulty analysis, and fit analysis. Details of these analyses are presented in Section 5 of Volume 1.

3.6 Development and Maintenance of the Item Pool

As described earlier, new items are developed each year to be added to the operational item pool after being field-tested. Several factors determine the development of new items. The item development team conducts a gap analysis for distributions of items across multiple dimensions, such as item counts, item types, item difficulty, and numbers in each reporting category.

In spring 2023, field-test items were embedded in online forms in grades 3 through 10 ELA Reading, grades 3 through 8 Mathematics, and Mathematics EOC tests. An independent field test for ELA Writing in grades 4-10 was also conducted in spring 2023. All assessments were CAT tests with a predetermined number and location of field-test items. Table 12 and Table 13 provide the number of field-test items by type for ELA Reading, Mathematics, and Mathematics EOC. Table 14 provides the number of writing prompts for each grade.

Table 12: Number of ELA Reading Field-Test Items by Type

Item Type	3	4	5	6	7	8	9	10
EBSR	17	42	49	31	31	33	33	22

Item Type	3	4	5	6	7	8	9	10
HT	1	4	2	3	6	2	12	7
MC	147	266	245	206	211	175	222	169
MI	11	11	19	14	20	8	15	9
MS	21	54	48	36	17	20	22	19
Two-Part HT	1	1	0	0	0	0	1	0

Table 13: Number of Mathematics and EOC Field-Test Items by Type

Item Type	3	4	5	6	7	8	Algebra 1	Geometry
EQ	99	71	122	120	57	26	36	67
ETIC	37	18	32	27	13	8	7	22
GI	0	0	1	0	0	1	6	10
HT	0	0	0	0	0	0	0	1
MC	98	68	128	85	67	32	23	47
MI	31	14	18	14	3	2	2	7
MS	36	25	20	18	5	6	1	5
Multi	2	2	4	4	8	0	9	10

Table 14: Number of ELA Writing Field-Test Prompts by Grade

Grade	Number of Prompts
4	10
5	10
6	11
7	10
8	10
9	12
10	16

3.7 Alignment Process for Existing Items and Results from Alignment Studies

A third-party, independent alignment study was conducted in February 2016. This report can be found in Volume 4, Appendix D, of the *Florida Standards Assessments 2015–2016 Technical Report*.

4. TEST CONSTRUCTION

4.1 OVERVIEW

During the 2022-2023 school year, the Florida Department of Education (FDOE) began transitioning from the fixed form Florida Standards Assessment (FSA) to the computer adaptive Florida Assessment of Student Thinking (FAST). In spring 2022, the first set of FAST items developed to align with the Benchmarks for Excellent Student Thinking (B.E.S.T.) standards were field-tested. In summer 2022, field test items were calibrated and placed on the Florida Standards Assessment (FSA) scale. Consistent with the PM1 and PM2 administrations, the spring 2023 FAST summative PM3 administration (as well as EOC Algebra 1 and Geometry) utilized CAI's adaptive algorithm to administer tests using these pre-equated items on the FSA scale. During this transition year, scores were reported to students on the FSA scale.

Subsequently, calibrations in summer 2023 placed items in ELA Reading at grades 3-10, and Mathematics at grades 3-8, on a common vertical FAST scale via a linking design. EOC Algebra 1 and Geometry were placed on the B.E.S.T. scale. Standard settings were conducted for all grades in ELA Reading, Mathematics, ELA Writing, Algebra 1, and Geometry. In the 2023–24 school year and beyond, FDOE will start reporting scores on the new FAST scale.

In addition to the online computer adaptive test, Florida also has accommodated forms. Accommodated forms were administered to students in lieu of the online forms if such a need was indicated on their Individualized Education Program (IEP) or Section 504 Plan. For the mathematic EOC assessments, Algebra 1, and Geometry, only one accommodated form was given. Accommodated forms used online parameters for scoring purposes and no calibrations were done on the accommodated forms.

4.2 ITEM SELECTION ALGORITHM

CAI's adaptive algorithm takes as input two sources of information: an item pool and a test blueprint. The adaptive algorithm is then configured to execute maximally adaptive test administrations under the constraint of blueprint match. Configuration of the adaptive algorithm is critical because the composition of the item pool, which changes from administration to administration, interacts with the blueprint to influence the performance of the adaptive algorithm.

Item Pool

CAI's ability to administer various state item pools is proven. For example, CAI administered items from the Smarter Balanced item bank during the 2013 pilot test and the 2014 field test. CAI designed and built the item renderers shared by the open-source version of the test delivery engine and CAI's own version of the item-rendering software. These renderers ensure that the items appear to students exactly as they did in the field test.

Test Blueprint

Test blueprints may contain specifications from the content hierarchy (strand, benchmark, standard, etc.) and other constraints, such as item type, or any other test item attribute that may be stored. CAI's adaptive engine supports blueprints that meet the following conditions (which have

been advocated by the Consortium for Citizens with Disabilities, an umbrella group encompassing most national advocacy groups for students with disabilities and other exceptional students):

1. Every student is tested on the full range of grade-level content, with no discernible differences in the content assessed.
2. Every student is tested on items measuring the same mix of cognitively complex skills, with no discernible difference—regardless of student proficiency.
3. Every student is tested on items reflecting the full range of other aspects of the grade-level curriculum as may be appropriate for the grade and subject.
4. Students are tested on items that provide the best measurement possible within these constraints.

These four principles ensure that every student can accurately demonstrate his or her academic skills and knowledge across the entire grade-level curriculum. CAI's adaptive algorithm supports blueprints that align with these principles.

Item Selection

The adaptive algorithm, built on our partnerships with client states over the years, ensures that each student will receive a test that (1) matches the blueprint and (2) contains the items that best match their performance level, as defined by the blueprint. To accomplish this goal, the algorithm implements a highly parameterized multiple-objective utility function that includes

- a measure of the content match to the blueprint,
- a measure of overall test information, and
- measures of test information for each reporting category on the test.

We define an objective function that measures an item's contribution to each of these objectives, weighting them to achieve the desired balance among them. The equation below sketches this objective function for a single item.

$$f_{ijt} = w_2 \left(\frac{\sum_{r=1}^{R} s_{rit} p_r d_{rj}}{\sum_{r=1}^{R} d_{rj}} \right) + w_1 \sum_{k=1}^{K} q_k h_{1k}(v_{kijt}, V_{kit}, t_k) + w_0 h_0(u_{ijt}, U_{it}, t_0)$$

Where the *w* terms represent user-supplied weights that assign relative importance to meeting each of the objectives, d_{rj} indicates whether item *j* has the blueprint-specified feature *r*, and p_r is the user-supplied priority weight for feature *r*. The term s_{rit} is an adaptive control parameter that is described below. In general, s_{rit} increases for features that have not met their designated minimum as the end of the test approaches.

The remainder of the terms represent an item's contribution to measurement precision:

- v_{kjit} is the value of item *j* toward reducing the measurement error for reporting category *k* for test taker *i* at time of selection *t*; and
- u_{ijt} is the value of item *j* in terms of reducing the overall measurement error for test taker *i* at time of selection *t*.

The terms U_{it} and V_{kit} represent the total information overall and on reporting category k, respectively.

The term q_k is a user-supplied priority weight associated with the precision of the score estimate for reporting category k. The t terms represent precision targets for the overall score (t_0) and each score reporting category score.

The functions h(.) are given by:

$$h_0(u_{ijt}, U_{it}, t_0) = \begin{cases} au_{ijt} \text{ if } U_{it} < t_0 \\ bu_{ijt} \text{ otherwise} \end{cases}$$

$$h_{1k}(v_{kijt}, V_{kit}, t_k) = \begin{cases} c_k v_{kijt} \text{ if } V_{kit} < t_k \\ d_k v_{kijt} \text{ otherwise} \end{cases}$$

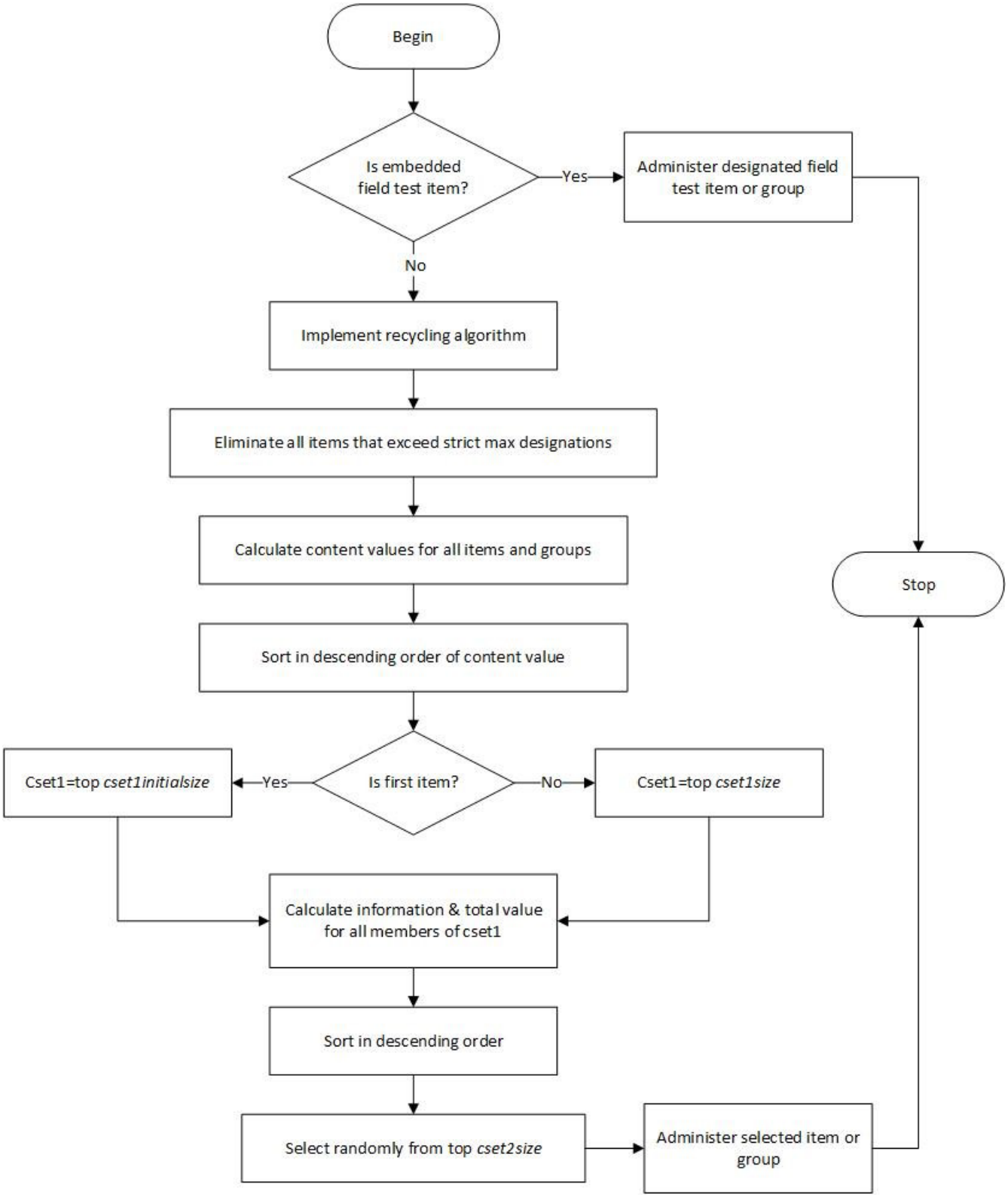
Items can be selected to maximize the value of this function. This objective function can be manipulated to produce a pure, standards-free adaptive algorithm by setting w_2 to zero or to produce a completely blueprint-driven test by setting $w_1 = w_0 = 0$. Adjusting the weights to optimize performance for a given item pool will enable users to maximize information subject to the constraint that the blueprint is virtually always met. We note that the computations of the content values and information values generate values on very different scales and that the scale of the content value varies as the test progresses. Therefore, we normalize both the information and content values before computing the value of the equation.

This normalization is given by $x = \begin{cases} 1 \; if \; min = max \\ \frac{v - min}{max - min} \; otherwise \end{cases}$, where min and max represent the minimum and maximum, respectively, of the metric computed over the current set of items or item groups.

Items (or groups of items in the case of the ELA tests) are sorted by their "content value," their value toward meeting the content constraints in the blueprint. Information measures are added to the content measures, and the items are sorted based on their overall value for the objective function. The final item selection is made based on a random selection from among the small subset of items that have the highest combined content and information value.

Figure 1 summarizes the item selection process. If the item position has been designated for a field-test item, then a field-test item is administered. Otherwise, the adaptive algorithm is triggered.

Figure 1: Item Selection Process



BLUEPRINT MATCH

Configuration of the adaptive algorithm for the spring 2023 administration was designed to administer tests meeting blueprint specifications while also maximizing test information to student ability for Mathematics tests. In the adaptive item-selection algorithm, item selection takes place in two discrete stages: blueprint satisfaction and match-to-ability. Due to the Operational Field Test design for the ELA Reading calibrations (for further details see Volume 1), the ELA Reading tests were only configured to meet the blueprint specifications.

While the simulation results described in the spring 2023 Simulation Summary Report (see Appendix F) indicated that the configuration resulted in the test administrations meeting all blueprint match requirements, it is also important to evaluate the blueprint match rate for the actual test administrations.

Appendix G, Spring 2023 Operational Item Blueprint Match, contains the operational item blueprint-match results for the spring 2023 grades 3–8 Mathematics, grades 3–10 ELA Reading and EOC Algebra 1 and Geometry. For the blueprint match analysis, only students who completed all parts of the test were included. As can be seen in Appendix G, in almost all assessments, all reporting categories met the blueprint or blueprint range. In addition to blue print match, the observed percentage of reading passage types by grade is documented in Table 15.

By 2024 we are aiming to achieve 100% blueprint match for all reporting categories and meet passage maximums for all grades by increasing size of pool and enable item recycling in the adaptive algorithm if necessary. Eventually, with a large enough item pool, item recycling will not be necessary.

Table 15: Observed Spring 2023 Percentage of ELA Reading Passage Types by Grade

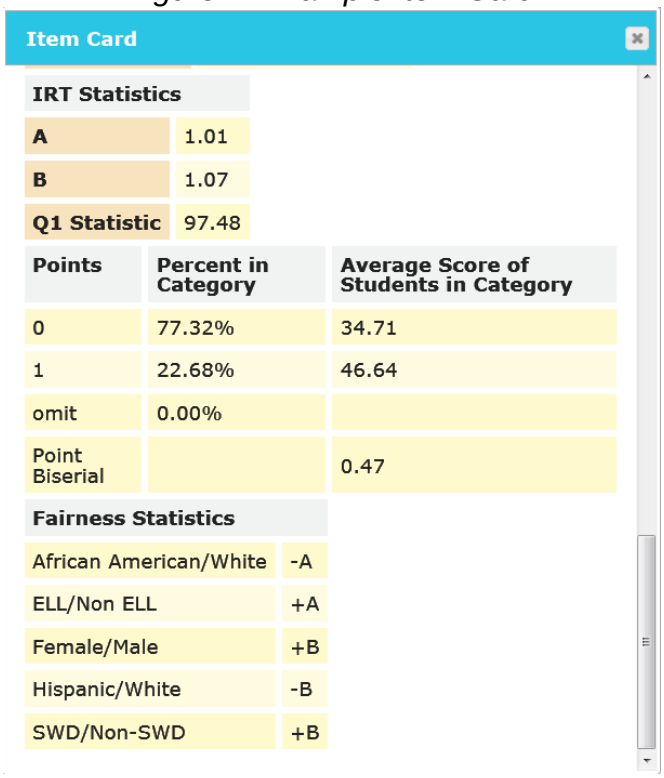
Grade	Informational	Literary
3	40%-60%	40%-60%
4	50%	50%
5	50%	50%
6	40%-50%	50%-60%
7	40%-60%	40%-60%
8	40%-50%	50%-60%
9	50%	50%
10	40%-60%	40%-60%

4.3 Test Construction Summary Materials

4.3.1 Item Cards

Item cards, generated within ITS, contained statistical information about an individual item. Item cards contained classical item statistics, IRT statistics, and DIF statistics. When possible, item cards also contained a screen capture of the item. This was not possible in the case of some technology-enhanced items. In these instances, the items were viewed directly in ITS. Item cards were typically used to determine the viability of an individual field-test item for operational use in the next administration. Figure 2 shows one example of an item card.

Figure 2: Example Item Card



4.3.2 Bookmaps for Accommodated Forms

Bookmaps were only provided for accommodated forms. A bookmap is a spreadsheet that lists the characteristics of all items on a form. Bookmaps contain information such as:

- Item ID
- Item position
- Form
- Grade
- Role (e.g., operational or field-test)
- Item format (e.g., MC)
- Point value

- Answer key
- Reporting category
- DOK

Bookmaps are used as an accessible resource by both content specialists and psychometricians to find information about a test form. Bookmaps differ from item cards in that there are no statistical summaries in a bookmap. Bookmaps contain useful information regarding the forms that are built in ITS.

4.4 ACCOMMODATION FORM CONSTRUCTION

Student scores should not depend upon the mode of administration or type of test form. Because the FAST grades 3–10 ELA Reading and grades 3–8 Mathematics tests were administered in an online test system, scores obtained via alternate modes of administration must be established as comparable to scores obtained through online testing. During test development, forms across all modes were required to adhere to the same test blueprints and content-level considerations. This section outlines the overall test development plans that ensured the comparability of online tests and accommodated tests.

To create the spring 2023 accommodated forms, CAI ran simulations in the summer of 2022 (based on the new FAST/B.E.S.T. blueprints and CAT algorithm) for each grade. Various simulation options were explored, examined, and discussed by psychometricians from FDOE and CAI. The final simulation setting for form creation was selected based on its overall outcome in terms of item pool size, blueprint constraints, and simulation results such as the item distribution report, overall test summary of violations, and standard error of ability estimates. For each subject, five forms were chosen from the online forms generated using the preferred simulation settings.

These five forms came from simulated test takers with their ability closest to the proficiency cuts. The test information function plots, test characteristic curves, and standard error of ability plots were compared. In these plots, particular attention was paid to the ability close to the proficiency cut. The forms with smaller standard error of ability estimates at the passing score were selected. As this was prior to the Spring 2023 calibrations and standard setting, the parameters and cuts were based on the pre-equated FSA scale.

Content reviewed the forms, and made any necessary item replacements taking into account suitability for inclusion in an accommodated form and psychometric feedback. To build accommodated forms, content specialists began with the selected forms and removed any technology-enhanced items that could not be rendered on accommodated forms or machine-scored. These items were then replaced with either MC items or other technology-enhanced items that could be rendered on accommodated forms from the same reporting category. In some instances, it was necessary to select replacement items from a different reporting category in order to satisfy statistical expectations; however, all parties ensured that each reporting category was still appropriately represented in the final test forms. Two of the forms with the best statistics were selected to be sent to FDOE for evaluation and selection of a final form.

REFERENCES

Allman, C. (2009). Position paper: *Making Tests Accessible for Students with Visual Impairments* (4th ed.). Louisville, KY: American Printing House for the Blind. Retrieved from https://sites.aph.org/wp-content/uploads/2017/09/Test-Access-Making-Tests-Accessible-2009.pdf

Shaftel, J., Benz, S., Boeth, E., Gahm, J., He, D., Loughran, J., Mellen, M., Meyer, E., Minor, E., & Overland, E. (2015). *Accessibility for Technology-Enhanced Assessments (ATEA) Report of Project Activities*. Research Report. University of Kansas.

Thompson, S. J., Johnstone, C. J., & Thurlow, M. L. (2002). *Universal Design Applied to Large-Scale Assessments* (NCEO Synthesis Report 44). Minneapolis: University of Minnesota, National Center on Educational Outcomes. Retrieved from https://nceo.umn.edu/docs/onlinepubs/synth44.pdf.

Webb, N. L. (2002). *Alignment Study in Language Arts, Mathematics, Science, and Social Studies of State Standards and Assessments for Four States*. Washington, DC: Council of Chief State School Officers.

Winter, P. (2010). Comparability and test variations. In P. Winter (Ed.), *Evaluating the Comparability of Scores from Achievement Test Variations* (pp. 1–11). Washington, DC: Council of Chief State School Officers.