



Statewide Assessments Guide

2023–2024

Updated April 2024



Table of Contents

1.0	Introduction.....	1
1.1.	FAST Overview	2
1.2.	B.E.S.T. EOCs and Writing Overview	3
1.3.	Science and Social Studies Assessments Overview	4
2.0	Test Content and Format	5
2.1.	Test Items.....	5
2.1.1.	Test Design	5
2.1.2.	Universal Design.....	6
3.0	Test Development.....	15
3.1.	Developing Items	16
3.1.1.	Item Writing.....	16
3.1.2.	Committee Reviews	16
3.1.3.	Field Testing	17
3.1.4.	Statistical Review.....	18
3.1.5.	Computer-Adaptive Testing.....	18
3.1.6.	Operational Test Items	21
3.1.7.	Released Test Items	27
4.0	Test Administration.....	28
4.1.	Schedules	28
4.2.	Computer-Based Testing (CBT).....	28
4.3.	Timing.....	29
4.4.	Personnel.....	29
4.5.	Test Security	30
4.6.	Administration Procedures.....	31
4.7.	Accommodations	33
4.8.	Test Invalidation	33
5.0	Scoring the Assessments.....	34
5.1.	Standard Setting	34
5.2.	Scoring Different Item Types	35
5.2.1.	Machine/Electronic Scoring.....	35
5.2.2.	Hybrid Scoring – B.E.S.T. Writing.....	36
5.2.3.	Reported Measures.....	39

5.2.4. IRT Scoring.....	40
6.0 Reporting Results.....	42
6.1. Reports	42
6.2. Know Your Data: Advanced Reports	43
6.3. Know Your Schools: Florida School, District, and State Report Cards	43
7.0 Glossary	44
8.0 Guide to Useful Resources.....	49
8.1. Writing Resources.....	49
8.2. Fact Sheets	49
8.3. Graduation Requirements	49
8.4. Sample Test Materials.....	49
8.5. Technical Reports	50
8.6. Test Administration Manuals.....	50
8.7. Test Design Summaries and Blueprints.....	50
8.8. Test Item Specifications.....	50
8.9. Schedules	51
8.10. Standard Setting	51
8.11. Understanding Reports.....	51
8.12. Useful Links.....	51
Appendix A: Statistical Indicators Used in Data Analysis.....	52
Indicator Definitions	52
Reliability Measures	53
Appendix B: History, Requirements, and Uses.....	55
History of Florida’s Statewide Assessments Program.....	55
Overview of State Statute and State Board of Education Rules Governing the Statewide Student Assessment Program.....	57
State-Required Uses of Statewide, Standardized Assessment Results	60
Primary Statutory Authority for the Statewide Student Assessment Program	60

1.0 Introduction

The primary purpose of Florida’s VPK–12 statewide assessments program is to measure students’ achievement of Florida’s educational standards. Assessment supports instruction and student learning. Assessment results help Florida’s educational leadership and stakeholders determine whether the goals of the education system are being met. Assessments help Florida determine whether we have equipped our students with the knowledge and skills they need to be ready for careers and college-level coursework.

Florida’s statewide assessments also provide the basis for student, school, and district accountability systems. Assessment results are used to determine school and district grades, which give citizens a standard way to determine the quality and progress of Florida’s education system. While assessment plays a key role in Florida’s education system, it is important to remember that testing is not an end in itself, but a means to an end. Florida’s assessment and accountability efforts have had a significant positive impact on student achievement over time.

Beginning with the 2022–23 school year, Florida’s statewide, standardized assessments in Reading, Writing, and Mathematics are aligned with the Benchmarks for Excellent Student Thinking (B.E.S.T), and assessments administered in Voluntary Prekindergarten (VPK) programs are aligned to the Florida Early and Developmental Standards (FELDS): Four Years Old to Kindergarten. The Florida Assessment of Student Thinking (FAST), which includes VPK Early Literacy, kindergarten through grade 10 ELA Reading, and kindergarten through grade 8 Mathematics assessments, will be administered as progress monitoring assessments, which students will participate in three times per year. B.E.S.T. assessments that are not part of the FAST progress monitoring program include grades 4–10 Writing and end-of-course (EOC) assessments in Algebra 1 and Geometry.

The *Statewide Assessments Guide* provides information about the grades 3–10 FAST assessments, B.E.S.T. assessments (Writing, Algebra 1 EOC, and Geometry EOC), Statewide Science Assessments, and Science and Social Studies EOC assessments in Biology 1, Civics, and U.S. History. This information includes the considerations governing item and test development, the mechanics of item and test scoring, and the meaning of the different assessment scores. Such an understanding can be useful for helping educators, parents, and students know more about the entire assessment process, including the time and resources that contribute to each phase from development to score reporting. Much of the information in this guide has appeared in other publications and on the Florida Department of Education (FDOE) website, and, although some of the information about the assessment program is technical, the guide is written for those without specialized knowledge of psychometrics or educational measurement. Technical information is presented at the conceptual level first, as well as in the context of its relevance to the tests. Appendix B provides an overview of the program’s history, as well as statutory requirements and uses.

Florida’s statewide assessments measure student achievement of the standards contained in the Florida state academic standards. Florida’s educational standards were developed with the goal of providing all students with an education based on high expectations. This guide contains information about the FAST, B.E.S.T., and Science and Social Studies assessments only; the

FDOE website contains information about the [Florida Alternate Assessment \(FAA\)](#), [ACCESS for ELLs](#), the [Florida Civic Literacy Examination \(FCLE\)](#), and the [National Assessment for Educational Progress \(NAEP\)](#).

Statewide assessment development is guided by the active involvement of Florida educators. FDOE maintains open communication with Florida educators regarding how the program and the various associated processes and activities might be improved. To ensure that the statewide assessments are accurate measures of Florida's academic standards, Florida educators are encouraged to become familiar with the process, remain up to date on new developments, and provide feedback via committee participation. This guide is intended to provide important background information, including further explanations of the role of educators in the assessment process.

For some Florida educators, parents, and other stakeholders, much of the information in this guide may be new; however, the development and implementation of the statewide assessments program have been shaped by the active involvement of thousands of Florida educators serving on various committees. Since 1995, educators have guided the development of Florida's educational standards, the determination of which standards to assess, how to assess the standards on statewide assessments, and how writing responses should be scored. In addition, all test items are reviewed and accepted by committees of Florida educators. Other committee participants include Florida citizens who share a stake in the education of Florida's children, as well as local and national experts in psychometrics.

This guide is produced in an online format only and contains links to resources for additional information throughout.

For information related to the information in the guide, please email Assessment@fldoe.org.

1.1. FAST Overview

In March 2022, Governor Ron DeSantis signed new legislation that replaced the grade-level Florida Standards Assessments (FSA) in English Language Arts (ELA) and Mathematics with the Coordinated Screening and Progress Monitoring System, also called the Florida Assessment of Student Thinking (FAST), required in section 1008.25(9), Florida Statutes (F.S.).

FAST assessments are aligned to the B.E.S.T. Standards. Because all FAST assessments are computer-adaptive, items become progressively harder as students successfully respond to items and easier if students answer more questions incorrectly, but in either scenario, the selected items measure the same knowledge and skills determined by the test blueprint. Each progress monitoring (PM) event covers the entire test blueprint for the full grade-level content. The assessments are administered three times during each school year: once at the beginning of the year (PM1), once in the middle of the year (PM2), and once at the end of the year (PM3). The ELA and Mathematics blueprints show the percentage of items in each reporting category students will encounter during each PM window. More information on computer-adaptive testing and test blueprints is found later in this guide.

The following are descriptions of each PM opportunity for students:

- PM1 – because this administration occurs at the very beginning of the school year, it is designed to provide a baseline score so teachers can track student progress in learning the B.E.S.T. Standards from PM1 to PM2.
- PM2 – by the time of this administration window, students will have had an opportunity to learn the grade-level standards, and this administration provides a mid-year score to compare to the baseline score from PM1.
- PM3 – this last administration provides a summative score that will accurately measure student mastery of the B.E.S.T. Standards at the end of the school year. While PM1 and PM2 are for informational purposes only, PM3 is used for school accountability.

Information in this guide related to FAST is focused on **grades 3–10** assessments only. Similar information for FAST at grades VPK–2 can be found in the following resources:

- [Research Foundation for Star Adaptive Assessments: Science of Star](#)
- [2023–2024 Kindergarten through Grade 2 FAST Fact Sheet](#)
- [Renaissance Star Sample Test Materials](#)
- [Star Parent Reports](#)
- [FAST K–2 Benchmark Reporting](#)

More detailed information about the FAST program is found throughout this guide. Additional resources can be found on the [FAST Portal](#).

1.2. B.E.S.T. EOCs and Writing Overview

In Winter 2022, EOC assessments aligned to the B.E.S.T. Standards in Algebra 1 and Geometry were administered for the first time. These tests are administered four times per year for students completing a respective course, or students who need to retake one of these assessments. Like the FAST assessments, these assessments are computer-adaptive, administered in one day, and scores are reported within 24 hours of a student’s test attempt.

In Spring 2023, the B.E.S.T. Writing field test was administered to a representative sample of Florida students in grades 4–10. The assessments are computer-based for all grade levels and consist of one text-based constructed-response item (i.e., students read a variety of texts and respond to a prompt). The rubrics used for the scoring of the Writing assessment are based on the B.E.S.T. ELA Standards. While the FSA Writing assessment contributed to the overall ELA score (combined with FSA Reading), the new B.E.S.T. Writing assessment is a standalone test that does not contribute to the FAST ELA Reading score.

B.E.S.T. Writing is administered once a year, with the first operational administration in Spring 2024.

1.3. Science and Social Studies Assessments Overview

Beginning in Spring 2024, the summative Statewide Science Assessment in grades 5 and 8, as well as the Biology 1, Civics, and U.S. History EOC assessments, were delivered in a computer-adaptive format that allows for immediate reporting. While the core content for these tests will not change, some administration details (e.g., reduced test length) and blueprint specifications (e.g., number of items each student will see) have been updated.

The Fall and Winter 2023 administrations of the science and social studies EOC assessments were computer-based, fixed form tests, and results were available for all students after the testing window as in previous years. For the purposes of this guide, information for the science and social studies EOC assessments will be presented to reflect the computer-adaptive Spring 2024 administration.

2.0 Test Content and Format

The statewide assessments are **criterion-referenced tests** that are intended to measure whether students have made progress on the B.E.S.T. ELA and Mathematics Standards and Florida’s Science and Social Studies Standards. Statewide assessments are developed to meet specific technical criteria and to ensure that all students have access to the test content via principles of universal design and appropriate accommodations.

Statewide assessments are delivered in a computer-based test (CBT) format with a paper-based test (PBT) format available as an accommodation for eligible students. The item types listed in this section vary based on subject and delivery mode.

Online Sample Test Materials for all subject tests are available on the [portal](#).

2.1. Test Items

When taking statewide assessments, all students of the same grade level, or subject area for EOC assessments, respond to a range of operational items on each test that count toward students’ scores. Because Florida’s assessments are computer-adaptive, these items are selected from a common item bank.

In the spring administrations, field-test items are also found on all students’ tests but do not count toward students’ scores. Field-test items are administered to students only to gather data on the items for future operational use.

The next three subsections provide additional information about the different content areas and detail the knowledge and skills assessed, item types, and other subject-specific information for each area.

2.1.1. Test Design

The Test Design Summaries list the range of operational items from each reporting category that is required on each test form. These documents help guide an algorithm for computer-adaptive item selection to ensure that reporting categories and standards are correctly represented on each test attempt.

The Test Item Specifications are based on Florida’s academic standards and course descriptions and provide detailed guidance for item writers and reviewers to ensure that items are aligned to the standards they were intended to measure.

More information about test design is provided in Section 3.0, Test Development.

2.1.2. Universal Design

The application of universal design principles helps the Department develop assessments that are accessible to the greatest number of test takers, including students with disabilities and English language learners. To support the goal of providing access to all students, the test maximizes readability and emphasizes compatibility with accommodations. Test development also includes a review for potential bias and sensitivity issues. FDOE trains both internal and external reviewers to revise test items, allowing for the widest possible range of student participation. Item writers must attend to the best practices suggested by universal design, including, but not limited to:

- reduction in wordiness,
- avoidance of ambiguity,
- selection of reader-friendly construction and terminology, and
- consistently applied concept names and graphic conventions.

Universal design principles also inform decisions about test layout and design, including, but not limited to, font size, line length, spacing, and graphics.

English Language Arts (ELA)

FAST ELA Reading

FAST ELA Reading is based on the English Language Arts B.E.S.T. Standards. FAST ELA Reading employs a wide variety of written material to assess students' reading comprehension as defined in the B.E.S.T. Standards. There are two types of reading passages: fiction and informational.

Informational passages provide readers with facts about a particular subject and may include magazine and newspaper articles, editorials, and biographies. Literary passages are written primarily for readers' enjoyment and may include short stories, poems, folk tales, and selections from novels. Some passage sets may also include multimedia items, such as an image or illustration. Most passages are selected from published sources, although some may be written expressly for the Florida Statewide Assessments. In addition, section 1008.22, F.S., requires that some of the reading passages for ELA assessments incorporate grade-level core curricula content from social studies. The range of words per passage increases across grade levels.

Table 1 provides an approximate word count range for a text or text set.

Table 1: Approximate Word Count Range for Text or Text Set

Grade	Range of Number of Words
3	100–700
4	100–900
5	200–1000
6	200–1100
7	300–1100
8	350–1200
9	350–1300
10	350–1350

Table 2 shows the percentages of FAST ELA Reading text types by grade band.

Table 2: Percentages of FAST ELA Reading Text Types

Grade Band	Fiction	Informational
3–10	50%	50%

For all grade levels tested, FAST ELA Reading assesses what students know and are able to do in the broad reporting categories listed in Table 3. More detailed explanations can be found in the [Achievement Level Descriptions](#).

Table 3: FAST ELA Reading Grades 3-10 Percentage of Points by Reporting Category

Reporting Category	Percentage of Assessment
Reading Prose and Poetry	25–35%
Reading Informational Text	25–35%
Reading Across Genres & Vocabulary	35–50%

The difficulty of the concepts assessed progresses systematically from grade to grade, as does the complexity of the texts presented to the student at each grade level. For more information about the specific benchmarks assessed within each reporting category at each grade level, please see the [Test Design Summary and Blueprint](#).

B.E.S.T. Writing

B.E.S.T. Writing is based on the English Language Arts B.E.S.T. Standards. B.E.S.T. Writing is administered once a year, each spring, to students in grades 4–10, and writing scores are reported separately from FAST ELA Reading. The writing tests are text-based, meaning students will read two to four passages related to the topic and then create a written response to a prompt using information from the provided passages to support their writing. There are two writing modes for all grade levels: argumentative and expository. Only one mode of writing is assessed in each grade each year.

There are three domains, each worth up to four score points, for B.E.S.T. Writing:

- Purpose/Structure
- Development
- Language

For information about which benchmarks are covered in B.E.S.T. Writing, please see the [Test Design Summary and Blueprint](#). For more information about the assessment, please see the [B.E.S.T. Writing Fact Sheet](#).

Mathematics

FAST Mathematics is based on the Mathematics B.E.S.T. Standards. FAST Mathematics is administered to students in grades 3–8, and two mathematics B.E.S.T. EOC assessments, Algebra 1 and Geometry, are administered to students enrolled in and completing the respective course (or an equivalent course). See the [B.E.S.T. Algebra 1 and Geometry End-of-Course Assessments Fact Sheet](#) for a list of courses.

Reference sheets and calculators are provided for certain assessments. More information about these resources can be found in the [Calculator and Reference Sheet Policies for Florida Statewide Assessments](#) document.

For all grade levels and subjects tested, FAST Mathematics and B.E.S.T. EOC assessments assess what students know and are able to do in the broad reporting categories listed in Table 4. More detailed explanations can be found in the [Achievement Level Descriptions](#).

Table 4: FAST Mathematics and B.E.S.T. EOC Percentage of Items by Reporting Category

Grade 3 Reporting Category	Percentage of Assessment
Number Sense and Additive Reasoning	23–29%
Number Sense and Multiplicative Reasoning	
Fractional Reasoning	
Geometric Reasoning, Measurement, and Data Analysis and Probability	
Grade 4 Reporting Category	Percentage of Assessment
Number Sense and Operations with Whole Numbers	31–37%
Number Sense and Operations with Fractions and Decimals	
Geometric Reasoning, Measurement, and Data Analysis and Probability	
Grade 5 Reporting Category	Percentage of Assessment
Number Sense and Operations with Whole Numbers	23–29%
Number Sense and Operations with Fractions and Decimals	
Algebraic Reasoning	
Geometric Reasoning, Measurement, and Data Analysis and Probability	

Grade 6 Reporting Category	Percentage of Assessment
Number Sense and Operations	33–42%
Algebraic Reasoning	25–36%
Geometric Reasoning, Data Analysis and Probability	25–36%
Grade 7 Reporting Category	Percentage of Assessment
Number Sense and Operations and Algebraic Reasoning	25–31%
Proportional Reasoning and Relationships	22–31%
Geometric Reasoning	22–28%
Data Analysis and Probability	22–28%
Grade 8 Reporting Category	Percentage of Assessment
Number Sense and Operations and Probability	22–28%
Algebraic Reasoning	22–28%
Linear Relationships, Data Analysis, and Functions	25–31%
Geometric Reasoning	22–28%
Algebra 1 Reporting Category	Percentage of Assessment
Expressions, Functions, and Data Analysis	31–38%
Linear Relationships	31–38%
Non-Linear Relationships	31–38%
Geometry Reporting Category	Percentage of Assessment
Logic, Relationships, and Theorems	33–40%
Congruence, Similarity, and Constructions	27–33%
Measurement and Coordinate Geometry	33–40%

Reading and Mathematics Item Types

All tests contain traditional multiple-choice items, and the following list contains all additional possible item types that may appear on Reading and Mathematics tests.

FAST ELA Reading Item Types

- 1. Multiple-Choice:** Students select one correct answer from four answer choices.
- 2. Selectable Text:** Excerpted sentences from the text are presented in this item type. These items may have one or two parts. In a two-part selectable text item, Part A might ask the student to make an analysis or an inference, and Part B might require the student to select the text to support the answer in Part A.
- 3. Multiselect:** The student is directed to select a specific number of correct answers from among the options provided. These items are different from multiple-choice items, which prompt the student to select only one correct answer.

- 4. Evidence-Based Selected Response (EBSR):** In this two-part item, the student is directed to select the correct answers from Part A and Part B. Typically, Part A is multiple-choice, whereas Part B may be either multiple-choice or multiselect. Part A often asks the student to make an analysis or an inference, and Part B requires the student to use the text to support the answer in Part A.
- 5. Table Match:** This item type presents options in columns and rows. Options may include words, phrases, sentences, quotations, line/paragraph/passage numbers, or images. The student is directed to select a box that matches a correct option from a column with a correct option from a row. Typically, there is only one correct option per row or column, though the number of correct answers may vary.
- 6. Multimedia:** Technology-enhanced content may include multimedia elements (e.g., images, charts, graphics, timelines). Audio elements are not included in grades 3-10. Multimedia elements may appear within passages (stimuli) or test items. Any of the item types described above may be used to assess the multimedia content. For accommodated paper-based assessments, multimedia content may be modified or replaced by paper-based items or stimuli that assess the same reporting category.
- 7. External Copy Interaction:** This item type includes a button allowing students to select text directly from the passage. When the student selects the external copy interaction button, the text of the associated stimulus will appear with highlighting. The student can then select the text that answers the question. These items may be used independently or as part of a two-part item. In a two-part item, Part A might ask the student to make an analysis or an inference and Part B might require the student to use the text to support the answer in Part A. In other cases, the two parts might be independent. This item type appears in computer-based assessments only.

FAST Mathematics and B.E.S.T. EOC Item Types

- 1. Multiple-Choice:** Students select one correct answer from four answer choices.
- 2. Editing Task Choice:** The student clicks a drop-down menu containing options to complete an equation or expression, a statement, or other component. The student then selects the correct response from the drop-down menu.
- 3. Selectable Hot Text:** The student is directed to click on one or more correct answers from among a number of options. When the student hovers over the options (e.g., phrases, sentences, numbers, or expressions), the text will highlight. This indicates that the text is selectable (“hot”). The options may be presented in various ways (e.g., as a list, embedded within text, or in a table). The student can then click on an option to select it.
- 4. Multiselect:** The student is directed to select all of the correct answers from among a number of options. These items are different from multiple-choice items, which allow the student to select only one correct answer.
- 5. Graphic Response Item Display (GRID):** The student uses the point, line, or arrow tools to create a response on a graph. The item type may also require the student to select numbers, words, phrases, or images and use the drag-and-drop feature to place them into a graphic.
- 6. Equation Editor:** The student enters a number, variable, expression, or equation, as appropriate to the test item, in a response box. The student is presented with a toolbar that includes a variety of mathematical symbols that can be used to create a response. The response box may be separate from the text of the item, or it may be embedded within text of the item (e.g., in line with a sentence or within a table).

7. **Matching Item:** The student checks a box to indicate whether information from a column header matches information from a row. The number of correct answer options per row or column may vary.

Any of the item types may be combined into a single item with multiple parts called a multi-interaction item. The student will interact with different item types within a single item. Each part could be a different item type.

Science

Florida’s Science assessments are based on the benchmarks found in the Science academic standards and the Florida Course Descriptions. The Statewide Science Assessment is administered to students in grades 5 and 8, and the Biology 1 EOC Assessment is administered to students enrolled in and completing the course (or an equivalent course). See the [Florida Science and Social Studies End-of-Course Assessments Fact Sheet](#) for a list of courses.

The Grades 5 and 8 Statewide Science Assessments include multiple-choice items only. Some items are part of a context-dependent (CD) set, where a longer stimulus is provided, such as a short article or description of an experiment, and students respond to several items associated with that stimulus. For both grade levels tested, the Statewide Science Assessment tests what students know and are able to do in the broad reporting categories listed below. More detailed explanations can be found in the [Grade 5 Science Achievement Level Descriptions](#) and [Grade 8 Science Achievement Level Descriptions](#).

Grade 5

- **Nature of Science**
Students evaluate investigations and experiments, organize data, identify the control group in an experiment, interpret data and analyze information, and distinguish between observations and opinions.
- **Earth and Space Science**
Students distinguish among objects in our solar system, identify categories of rocks and characteristics of minerals, differentiate between physical weathering and erosion, identify characteristics associated with different climate zones, and identify factors that affect weather.
- **Physical Science**
Students identify basic forms of energy; identify familiar forces; trace the conversion of electric energy into other forms of energy; and distinguish relationships among mass, force, and motion.
- **Life Science**
Students identify the function of human body organs, compare life cycles of Florida plants and animals, identify adaptations in animals and plants that allow them to survive, and trace energy through a food chain.

Grade 8

- Nature of Science**
 Students identify test and outcome variables in an experiment, differentiate between experiments and investigations, analyze information to make inferences or predictions, differentiate between replication and repetition, and distinguish between theories and laws.
- Earth and Space Science**
 Students relate the positions of the Sun, Moon, and Earth that result in tides, moon phases, and eclipses; identify Earth changes due to weathering, erosion, and plate tectonics; and recognize that the Sun’s energy influences global atmospheric patterns.
- Physical Science**
 Students classify substances by physical properties, differentiate between physical and chemical change, distinguish between kinetic and potential energy, and differentiate contact forces and forces acting at a distance.
- Life Science**
 Students identify functions of the human body systems, classify organisms, identify ways genetic variation contributes to the scientific theory of evolution, determine probabilities for genotypic and phenotypic combinations, and distinguish relationships among organisms in a food web.

Table 1: Statewide Science Percentage of Items by Reporting Category

Grade	Nature of Science	Earth and Space Science	Physical Science	Life Science
5	17%	29%	29%	25%
8	19%	27%	27%	27%

Biology 1

The Biology 1 EOC Assessment measures student achievement of the state academic standards in science, as outlined in the Biology 1 course description. The test consists of multiple-choice items that measure what students know and are able to do in the broad reporting categories listed below. More detailed explanations can be found in the Biology 1 [Achievement Level Descriptions](#).

- Molecular and Cellular Biology**
 Students compare prokaryotic and eukaryotic cells, differentiate between mitosis and meiosis, relate the structure and function of the four major categories of biological macromolecules, and differentiate the processes of photosynthesis and cellular respiration.
- Classification, Heredity, and Evolution**
 Students identify evidence that supports the scientific theory of evolution, classify organisms into domains or kingdoms, identify scientific explanations of the origin of life, determine conditions required for natural selection, and analyze patterns of inheritance.

- **Organisms, Populations, and Ecosystems**

Students relate structure and function of organs and tissues in plants and animals; identify the structures and functions of organs in the human reproductive system, vascular system, central nervous system, and immune system; evaluate factors contributing to changes in population size; determine consequences of the loss of biodiversity; and evaluate the impact of biotechnology.

Table 2: Biology 1 Percentage of Items by Reporting Category

Category	Percentage
Molecular and Cellular Biology	35%
Classification, Heredity, and Evolution	25%
Organisms, Populations, Ecosystems	40%

Social Studies

Florida’s Social Studies assessments are based on the benchmarks found in the state academic standards and the Florida Course Descriptions. The Civics and U.S. History EOC assessments are administered to students enrolled in and completing the respective course (or an equivalent course). See the [Florida Science and Social Studies End-of-Course Assessments Fact Sheet](#) for a list of courses.

Civics

The Civics EOC Assessment measures student achievement of the state academic standards in Social Studies, as outlined in the M/J Civics course descriptions. The test consists of multiple-choice items that measure what students know and are able to do in the broad reporting categories listed below. More detailed explanations can be found in the Civics [Achievement Level Descriptions](#).

- **Origins and Purposes of Law and Government**

Students identify the origin, course, and development of the American legal and political traditions; the inherent conflicts involved in formulating those documents which would establish the nation; and how those concepts of the rule of law, limited government, and checks and balances remained constant through the first centuries of American history.

- **Roles, Rights, and Responsibilities of Citizens**

Students understand and define the concepts of citizen and citizenship with their corresponding obligations, rights, and responsibilities; explain the role of the Constitution in safeguarding individual rights and limiting government power; and evaluate the impact of relevant constitutional amendments and the significance and outcomes of landmark Supreme Court decisions.

- **Government Policies and Political Processes**

Students identify current political parties and formulate ideas regarding government, examine the impact of interest groups, evaluate political candidates, analyze the role of media in policy issues, identify appropriate government agencies for resolving policy debates, comprehend and differentiate concepts related to U.S. domestic and foreign policy, and describe how the United States has dealt with international conflicts.

- **Organizations and Functions of Government**

Students compare the different forms and systems of government, understand the role of the three branches of government, recognize the division of federal and state obligations and powers, articulate the constitutional amendment process, understand the judicial process, and compare the Constitutions of the United States and Florida.

Table 3: Civics Percentage of Items by Reporting Category

Category	Percentage
Origins and Purposes of Law and Government	25-30%
Roles, Rights, and Responsibilities of Citizens	15-20%
Government Policies and Political Processes	20-25%
Organization and Function of Government	20-25%

U.S. History

The U.S. History EOC Assessment measures student achievement of the state academic standards in social studies, as outlined in the United States History course descriptions. The test consists of multiple-choice items that measure what students know and are able to do in the broad reporting categories listed below. More detailed explanations can be found in the U.S. History [Achievement Level Descriptions](#).

- **Late Nineteenth and Early Twentieth Centuries (1860–1910)**

Students understand and articulate the impact of issues related to the Civil War, Reconstruction, the closing of the frontier, the industrialization of the nation, and changes in American society at the beginning of the twentieth century.

- **Global Military, Political, and Economic Challenges (1890–1940)**

Students understand and articulate the impact of the issues related to the rise of American military power; America’s increased involvement in world affairs; and changing social, political, and economic forces affecting the 1920s and 1930s.

- **The United States and the Defense of the International Peace (1940–2010)**

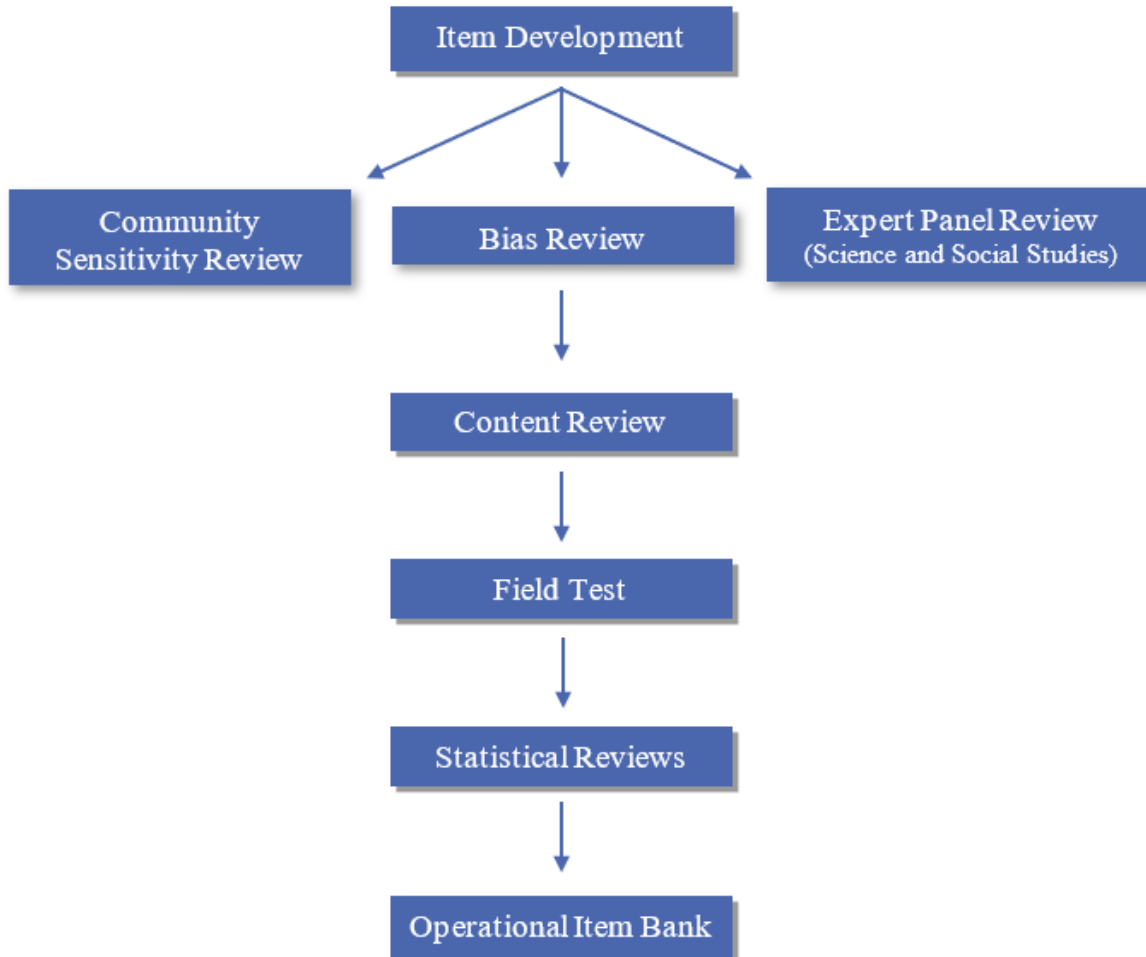
Students understand and articulate the impact of issues related to World War II, the Cold War, the social revolutions of the late twentieth century, and the challenges of the early twenty-first century.

Table 4: U.S. History Percentage of Items by Reporting Category

Category	Percentage
Late Nineteenth and Early Twentieth Century, 1860–1910	33%
Global Military, Political, and Economic Challenges, 1890–1940	34%
The United States and the Defense of the International Peace, 1940–Present	33%

3.0 Test Development

Developing statewide assessments to accurately measure achievement and accurately compare end-of-year results from one year to the next requires an extensive process involving many people with varied expertise. This process is overseen by the Florida Department of Education and annually integrates the work of FDOE’s Test Development Center (TDC), outside contractors, and several hundred Florida educators and citizens. This chapter provides details about each step in this process.



Before reading about the development process, it is important to understand two key concepts. The first relates to field testing items. When an item first appears on an assessment, it is as a field-test item and does not count toward a student’s score. After field testing, if the item is statistically sound, then it may be used on a test as an operational item, which counts toward a student’s score.

The second key concept relates to the nature of the item writing process. Item writers do not write a complete test in any given year. Instead, they write individual items that will go through a series of reviews. If items are accepted and have passed through each review successfully, then

the items become part of the item bank. The item bank is a database of items serving as the source from which the computer-adaptive algorithm selects appropriate items for each student.

3.1. Developing Items

There are five key stages in the development of an assessment item, from item writing through inclusion on an assessment as an operational item:

1. Item Writing
2. Committee Reviews
3. Field Testing
4. Statistical Review
5. Operational Testing

3.1.1. Item Writing

For each subject and grade level, criteria for item development are specified in the FAST, B.E.S.T., and Science/Social Studies Test Item Specifications. The specifications include the specific benchmarks, types of items used, guidelines for the relative balance of topics, item formats, cognitive complexity levels, and general guidelines to minimize non-content influences, such as confusing wording or poor graphics.

Each set of specifications is developed by FDOE in each of the four content areas. The specifications may be revised periodically to provide new sample items and reading passages.

Each year, for all grades and subjects, FDOE and the contractor agree on a list of benchmarks and item types for which items need to be written. This decision is based on a comparison of the benchmarks in the specifications with items already in the item bank. Then teams of item writers use the specifications to write new items for the designated benchmarks.

Item writers have teaching experience with varied and often specialized backgrounds and abilities. All item writers are required to attend a training session that includes a review of item specifications, cognitive complexity levels, characteristics for each item type, examples of good and poor items, scoring criteria, universal design principles, and an explanation of bias and community sensitivity concerns. After training, item writers are assigned to write and submit items for review. Items are reviewed and edited several times by FDOE content experts and the contractor before going on to the next stage of development.

3.1.2. Committee Reviews

Potential field-test items must be reviewed by several committees and FDOE before being approved for field testing with Florida students.

All items are reviewed by Bias Review Committees, composed of educators from Florida school districts and universities. In addition to some returning members, new committee members are invited to participate each year on an ad hoc basis. The purpose of a Bias Committee review is to look for any items, prompts, graphics, or passages that might provide an advantage or

disadvantage (unrelated to what is being measured) to a student with certain personal characteristics, such as those related to race or ethnicity, for example.

Similar to the Bias Review Committees, Community Sensitivity Committees are made up of Florida citizens associated with a variety of organizations and institutions. Membership is drawn from statewide parent organizations, community-based organizations, religious organizations, cultural groups, school boards, school district advisory councils, and business and industry from across the state. Reviewers are asked to consider whether the subject matter and language of test items, writing prompts, graphics, or reading passages will be acceptable to students, their parents, and other members of Florida communities. Issues of community sensitivity are distinct from bias because community sensitivity issues do not necessarily affect student success on an item, whereas bias may. Sensitive topics for Florida students may include hurricanes, for example, or other topics that may be considered offensive or too sensitive for students or that may distract students from the task at hand.

After each review committee meeting, a list of all members' comments is compiled and presented to FDOE for evaluation and inclusion in the materials used during the Item Content Review Committees that follow.

Item and Passage Content Review Committee members are Florida educators, including teachers and administrators from the targeted grade levels and subject areas, and school and district specialists from the content areas. Committee members determine whether the passages, graphics, and items are appropriate for the proposed grade levels. Committee members evaluate whether the items measure the standards, are clearly worded, have only one correct answer, and are grade-level or course appropriate. Committee members also recommend approval, modification, or rejection of the passages or items presented by FDOE. Additionally, Science and Social Studies items are reviewed by Expert Review Committees, panels of university-level and practicing research scientists, science-related industry experts, and legal experts. This review ensures the content accuracy of the test items in fields where information may change over time.

Following committee reviews, the passages and items go through a final editorial review process. Approved items are ready to enter the field-testing stage.

3.1.3. Field Testing

Field-test passages and items are embedded among the operational items in ELA Reading, Mathematics, Science, and Social Studies. For computer-adaptive assessments, each student will see approximately five field test items each spring.

In April 2023, grades 4–10 students in selected schools in the state participated in a standalone B.E.S.T. Writing Field Test to create an item bank for use in operational B.E.S.T. Writing, first administered in spring 2024. A standalone field-test administration is required to gather important data and establish scoring guidelines for the field-tested prompts, should they be used on future operational assessments. For B.E.S.T. Writing, Rangefinder Committees examine a representative set of student responses from writing field tests in order to establish scoring guidelines. Approximately 200 student responses representing a full range of possible score points are reviewed and committee members identify student responses reflective of each

specific point within each domain. The responses scored by the Rangefinder Committees are developed into materials for training teams of professional scorers. Committee members are Florida educators, including teachers from the targeted grade level and subject area, and school, district, and university specialists from the curriculum area. Before these prompts are used on a test to contribute to a student's score, the training materials created by the Rangefinder Committees are reviewed by a Rangefinder Review Committee.

Responses to field-test items do not count toward students' scores, and scores are not reported for a standalone writing field test. Students' responses to these items yield statistics that further reveal the quality of the item. Based on the analyses of field-test data, items are either rejected or placed in the item bank for use as operational items. After being accepted into the item bank, but before being used as operational items, writing prompts must undergo a further review.

Rubric Validation Committees review all responses to certain field-tested technology enhanced items for ELA Reading and Mathematics to determine whether all possible correct answers have been included in the scoring key. For instance, in an equation-editing task, a student could enter an expression as " $2x + 5$ " or " $5 + 2x$ " and the scoring system must account for all correct possibilities. Based on the input of the committees, FDOE establishes rules for how each item will be scored. The committees are comprised of Florida educators, including teachers from the targeted grade levels and subject areas and school and district curriculum specialists. These committees meet after each spring administration before field-test items are scored for statistical analyses.

3.1.4. Statistical Review

After field-test items have been scored, information about each item is electronically filed in the item bank. This information includes the item itself and the item statistics. Prior to being selected for inclusion as an operational item, the field-test statistics for the item must satisfy quality criteria. See Characteristics of Test Items in Section 3.1.6. for more detailed information about these criteria.

3.1.5. Computer-Adaptive Testing

A Computer-Adaptive Test (CAT) is a form of computer-based test that adapts to a student's ability level. In a CAT, each test question is selected by a computer algorithm that meets the published test blueprint addressing the content standards and is based on a student's performance on previously administered items within a test attempt. A CAT is also known as a tailored test since it adjusts the difficulty of the test items for a student's skill level, considering whether the student responded successfully to the preceding item or sets of items. Each time a student answers a question, his or her response helps determine the next question or set of questions that will be presented to the student. Computer-adaptive tests have long been used across the country and around the world, including in PreK–12 schools, postsecondary institutions, and in a wide variety of licensure and certification examinations. Results of computer-adaptive assessments have been proven to be valid and reliable measures of what test takers know and can do.

Item Selection Algorithm

Because Florida's assessments are computer-adaptive, traditional test construction is not needed as it is with fixed-form tests. However, before operational testing takes place, numerous simulation studies are conducted, considering the requirements of test blueprint, item bank features, and characteristics of student populations in Florida. Several follow-up statistical analyses are run to evaluate the performance of the item selection algorithm to ensure that it works correctly in terms of providing full content coverage, controlling item exposure rates, and maximizing score precision.

The adaptive algorithm uses three sources of information when selecting items for students: the item bank, the test blueprint, and student performance. The adaptive algorithm is then applied to select and administer items under the constraints of blueprint match and score precision. Configuration of the adaptive algorithm is critical because the composition of the item bank, which changes from administration to administration, interacts with the blueprint to influence the performance of the adaptive algorithm.

Item Bank

The test item banks contain items that address the breadth of Florida's standards at each subject and grade level and cover a range of item difficulties. This ensures two things: first, there are enough different items to ensure test blueprint coverage for each assessment attempt, and second, the varying difficulty levels allow for each test attempt to be tailored to a student's ability. As more items are developed and field tested, the item banks become more robust over time.

Test Blueprint

The test blueprints show the percentage of items in each reporting category students will encounter for each test attempt. For FAST, students will see the full scope of grade/course-level content that falls within the percentage of the overall test length for each PM attempt. Blueprint coverage ensures that each student sees a certain percentage of items from each reporting category; they will not necessarily see an item for every benchmark.

Using the blueprints, the algorithm ensures that the following constraints are met for each test:

1. Students are tested on the full range of grade-level content, with no discernible differences in the content assessed.
2. Students are tested on items reflecting the full range of other aspects of the grade-level curriculum as may be appropriate for the grade and subject.
3. Students are tested on items that provide the best measurement possible within these constraints.

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

Item Selection

The adaptive algorithm ensures that each student will receive a test that matches the blueprint and contains the items that best match his or her 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.

The algorithm defines an objective function that measures an item’s contribution to each of these objectives, weighting them to achieve the desired balance among them. Items (or groups of items in the case of the ELA Reading tests) are sorted by their “content value,” their value toward meeting the content constraints in the respective 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 subset of items that have the highest combined content and information value. When a test begins, students are presented with the first item of the test. The algorithm assigns a question that matches the test blueprint and is of medium difficulty, or, for FAST assessments, is based on a student’s performance on a previous test attempt.

The algorithm functions slightly differently depending on the test being administered:

- **Mathematics:** Adaptive at the item level.
 - Based on a student’s response to an item and blueprint requirements, the algorithm will select the next appropriate item.
- **Reading:** Adaptive at the passage/passage set level.
 - Based on a student’s responses to a passage set and blueprint requirements, the algorithm will select the next passage/passage set.
 - Once the algorithm selects the next set, it is locked in (regardless of whether students go back and change answers within the current set).
- **Science and Social Studies (Spring 2024):** Adaptive at the test blueprint level.
 - Each item a student receives will be randomly selected to meet blueprint requirements and cover a range of item difficulty levels.

Beginning in Spring 2025 and beyond, the tests will be fully adaptive at the item level, similar to mathematics tests.

A [Frequently Asked Questions \(FAQ\)](#) document is available on the FDOE website to address common questions regarding how Florida’s computer-adaptive assessments function. A [brief video](#) explaining the concepts described in this section is also available on the FDOE website and is suitable for teachers and families to help users better understand computer-adaptive tests.

3.1.6. Operational Test Items

An item becomes “operational” when it is administered in Florida public schools and is scored to contribute towards a student’s overall scale score. During the scoring process, FDOE reviews statistical data from student performance on operational items using many of the same statistical criteria as were used in the reviews of field-test items. Reviews ensure that the items meet established design and psychometric criteria, as indicated by the field-test results.

The next sections present descriptions of the desired characteristics of test items, as well as the measures taken to ensure items meet these criteria. Each section provides a general description of related characteristics, processes, and quality control measures. More detailed information on the statistical indicators and processes can be found in Appendix A.

Characteristics of Test Items

This section explains the various analyses performed on field-tested items to decide whether they will be used on a test. The statistical analyses described in this section are performed both after they are field tested and again after each time they appear as operational items to verify that the items performed as expected. Definitions for the terms throughout this section can be found in Appendix A.

Content Validity – Connection to a Benchmark

All test items must address a specific benchmark. Items are reviewed and evaluated for how well they address the benchmarks for which they were developed.

Quality Assurance Measures—Ensuring that items are written to specific standards is the responsibility of item writers, Item Content Review Committees, and FDOE. In fact, content validity is not quantifiable by the statistical analyses routinely performed in data analysis; however, item writers are given clear instructions about writing items to assess specific standards, and they are reviewed for direct connections to standards at several points in the development process.

Difficulty Level

Items that are very easy or very hard may provide useful information for some, but not all, students. For the majority of test takers, test items of moderate difficulty tend to provide the most information about what students know and can do. A moderately difficult item is not so easy that virtually all students answer it correctly, nor so difficult that virtually all students answer it incorrectly.

Quality Assurance Measures—After items have been written, but before they have been field tested, they are reviewed for grade-level and course appropriateness by FDOE and the Item Content Review or Prompt Review Committees.

After field testing, statistical analyses of student performance are used to verify that items are within an acceptable range of difficulty. For single-point items, one indicator of difficulty is the

p -value, an item's difficulty index expressed as the proportion of students who responded correctly (successfully) to an item. For multi-point items, one indicator of difficulty is the ratio of an item's mean score to the number of points possible (similar to the p -value). There is another index for item difficulty based on Item Response Theory (IRT) framework, which is a mathematical model that specifies the relation of student ability levels and a student's item response (Embretson & Reise, 2000). The b -parameter estimate of the item characteristic is the indicator of item difficulty. If an item falls outside the range of acceptable values for any of these, it may be rejected for further use.

Item Discrimination (Item-Test Correlation)

For an item to be useful on a test, there must be a positive correlation between students' success on an item and their success on the test as a whole. In other words, students who succeed on a given item should exhibit greater success on the test as a whole than students who do not succeed on that item. Similarly, students with relatively higher achievement on the test as a whole should exhibit greater success on any given item than students with relatively lower achievement. This relationship may seem obvious since the test score is based on the scores of individual items; however, among items there will be variation in the strength of the relationship with some items exhibiting only a minimal correlation. In rare cases, there may even be a negative correlation, meaning that students who succeed on an item exhibit lower levels of overall achievement on the test. Items with minimal or negative correlations with overall test success may be poorly worded, may have two correct answers, may not actually test what they are intended to test, or may assess something that is unrelated to what the other items test. Items that exhibit such characteristics are excluded from scoring and either revised for further field testing or removed from consideration for operational use.

Quality Assurance Measures—Using detailed item development guidelines and field testing is intended to reduce the number of items with low or negative item-test correlations. These guidelines and the multi-step process of item development usually result in well-written items that assess what they are intended to assess and that are aligned with the overall content of the test. As verification, however, point biserial correlation for single-point items and poly-serial correlations for multi-point items are generated and reviewed after both field testing and operational testing. A similar review is also conducted via IRT a -parameter estimates, as an indicator of item discrimination. Appendix A describes the statistical indices used to analyze test data.

Guessing

On a multiple-choice item with four choices, the likelihood of choosing the correct answer simply by guessing is about 25%. If the distractors (the incorrect alternative choices) are ineffective and most students are able to easily eliminate one or more of them and then select their answer from the remaining choices by guessing, the likelihood of guessing the correct answer increases. Choosing from among a reduced number of options as described here is called *pseudo-guessing*.

Instead of a four-choice item, the item essentially becomes a three- or two-choice item. To minimize pseudo-guessing on a multiple-choice item, item writers and reviewers are instructed to design items that include plausible distractors but have only one correct answer.

Quality Assurance Measures—After field testing, test developers examine data for each item, including the percent of students choosing each possible response and the IRT c -parameter estimates, as an indicator of pseudo-guessing. Items with unusually high pseudo-guessing indices or high c -parameters are rejected.

Freedom from Bias

An item is considered biased if it places a group or groups of students at a relative advantage or disadvantage due to characteristics, experiences, interests, or opportunities common to the group and unrelated to academic achievement (or irrelevant to the construct being measured).

Quality Assurance Measures—Instructions to item writers and reviewers call attention to the possibility of bias and include a checklist to ensure that items are free from bias. Two additional measures identify and eliminate potential bias.

First, items are reviewed by the Bias Review Committees, which note any potential bias and give their comments to item reviewers. In some cases, items are eliminated from further consideration at this point.

Second, in addition to the thorough reviews by the Bias and Community Sensitivity Review Committees, bias is also investigated via statistical analyses, called Differential Item Functioning (DIF). Items with DIF exhibit differences in probability of correct response between student groups, even though these groups have the same level of abilities on the subject being measured. DIF statistics are calculated for sex (male versus female) and ethnic groups (e.g., White versus Hispanic), disability status (students with disabilities versus others), and language status (English language learners versus others). These statistics not only allow FDOE to identify potentially biased items but also to understand the likely impact of the bias on student performance. Field-tested items can be rejected for future use as operational items based on these analyses.

Universal Design Principles

Applying universal design principles to the development of test questions yields assessments that are usable by the greatest number of students, including those with disabilities and non-native speakers of English. To support the goal of providing access to all students, the test maximizes readability, legibility, and compatibility with accommodations.

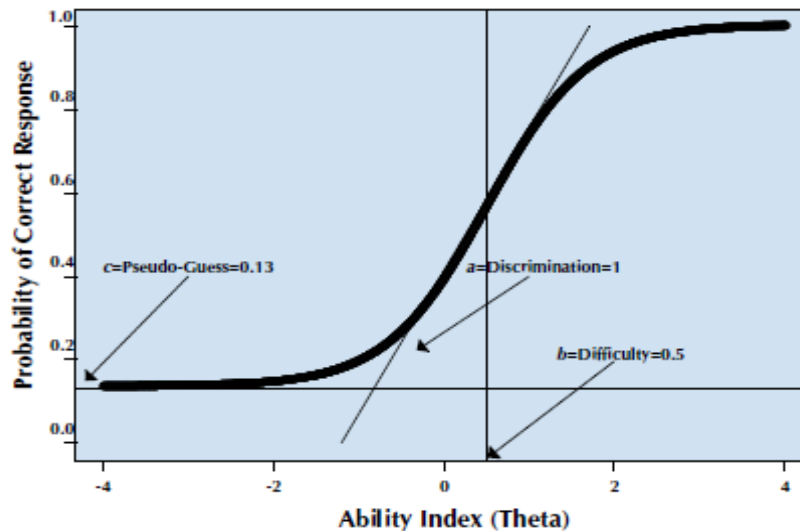
Quality Assurance Measures—FDOE trains both internal and external reviewers to write or revise items in such a way as to allow for the widest possible range of student participation. Item writers attend to the best practices suggested by universal design, including, but not limited to, reduction of wordiness, avoidance of ambiguity, selection of reader-friendly constructions and terminology, and application of consistently applied concept names and graphic conventions.

Universal design principles are also used to make decisions about test layout and design, including, but not limited to, type size, line length, spacing, and graphics. FDOE and the test contractor use the Test Production Specifications to ensure that test forms (on computer and paper) meet established high-quality standards. For test security reasons, the Test Production Specifications are not released to the public.

IRT Framework

The purpose of this section is to provide a broad summary of the statistical model used to score the assessments. Readers interested in more detailed information should consult the cited references as well as Appendix A. Scoring of the statewide assessments is built on IRT. Essentially, IRT assumes that test-item responses by students are the result of underlying levels of knowledge and skills, known as ability, possessed by those students and shown in the item characteristics. Items that fit the IRT model will have lower probabilities of correct responses from low-achieving students and higher probabilities of correct responses from high-achieving students.

In IRT analysis, a computer program creates a function for each item so that the resulting item characteristic curve most closely resembles the actual pattern of student responses. In this function, students' probability of success on an item corresponds to true levels of ability. For a multiple-choice item, the function incorporates three characteristics of the item: the a -, b -, and c -parameters. The a -parameter reflects the item's ability to distinguish between students above and below a given level; the b -parameter represents the relative difficulty of the item; and the c -parameter reflects the likelihood of low-achieving students guessing the correct answer. These parameters are estimated after items are field tested and reviewed during the data review meetings, where item parameters are carefully reviewed to determine if an item is suitable to become an operational item.



a = a function of the slope at point of inflection of the item characteristic curve
 b = theta value at point of inflection of the item characteristic curve
 c = lowest probability value of item characteristic curve

Items differ in their difficulty such that the position of the point of inflection of this curve (the vertical line on the example above) is higher or lower (to the right or to the left) along the theta (ability) scale. For example, the point of inflection of the item characteristic curve shown in the example is centered at about one-half a standard deviation above the zero point. An efficient test is composed of items with characteristic curves similar to this example, but with varying difficulties (points of inflection) that are positioned along the entire theta, or ability, scale. The three-parameter logistic (3PL) model (Lord & Novick, 1968)¹ is used to analyze multiple-choice items. The two-parameter logistic (2PL) model is a restricted version of the 3PL model where the c -parameter (pseudo-guessing) is set to zero. This model is used to analyze some single-point (correct or incorrect) items. The Generalized Partial Credit Model (Muraki, 1997)² is used to analyze multi-point items, such as writing prompts or technology enhanced items (e.g., open response items).

IRT item parameter estimates for all items on a test provide the means for determining scores for individual students. Because the item parameter estimates represent response probabilities, each student's achievement is assigned as the score most likely to correspond to that student's responses. In other words, student responses to a set of items and item parameter estimates are presented to an algorithm to derive the scores. The algorithm employs numerical methods to find the maximum likely score estimate for each student, based on their responses and item characteristics.

Quality Assurance Measures—The Technical Advisory Committee (TAC) is composed of 10–15 professionals with expertise in psychometrics and/or assessment. The members include Florida District Assessment Coordinators, Florida university faculty members, and nationally and internationally known assessment expert representatives of universities and state agencies outside Florida. In addition, the psychometric advisors of FDOE's contractor participate in the committee meetings. Committee members assist the FDOE by reviewing technical decisions and documents and by providing advice regarding the approaches FDOE should use to analyze and report assessment data. This committee meets twice a year.

Item Fit to the IRT Model

Data analyses conducted after field testing generate IRT statistics for each item. These statistics inform the degree to which the item differentiates between students of different abilities (the a -parameter), the difficulty of the item (the b -parameter), and the likelihood of success by pseudo-guessing (the c -parameter). The statistics are used to establish a mathematical model that specifies the relation of student ability levels and a student's item response. This mathematical model, also named Item Response Function, is expected to fit the observed pattern of student responses.

¹ Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

² Muraki, E. (1997). A generalized partial credit model: application of an EM algorithm. *Applied Measurement*, 7, 159–176.

Quality Assurance Measures—For each item, statistics describing the quality of fit to the model are generated. These statistics are derived by estimating the expected student performance on the item and then comparing this estimate to actual student performance on the item. For assessment data, there are established standards for values that indicate a good fit of the model.

Test Reliability

Statewide assessment scores are estimates of students' levels of achievement. A reliable score provides an accurate estimate of a student's true achievement. As with any estimate, there is some degree of measurement error. On a reliable test, the amount of error will be small. When there are sufficient numbers of test items in the item bank that reflect the intended content, are free from bias, are well-written, represent a range of difficulty, and have positive correlations to success on the test, the likelihood of the test being reliable will be high and the amount of error will be low. The measurement error is further reduced with a computer-adaptive test because the item selection algorithm employed by the CAT tailors test forms appropriate to each student's ability level.

Quality Assurance Measures—After the field-test administration is completed, IRT statistics are generated through the calibration process, the item bank is fed with recently field-tested items, simulation studies are conducted (before operational testing takes place), and several follow-up statistical analyses are run to evaluate the item selection algorithm's performance in terms of maximizing score precision, including expected IRT marginal reliability and expected Conditional Standard Error of Measurement (CSEM). If necessary, simulation conditions are modified to improve the expected score precision indices. After operational testing is completed, score accuracy is re-evaluated by means of observed IRT marginal reliabilities, observed CSEM, and Cronbach's alpha coefficients.

Test Fit to the IRT Model

The IRT model used in test development and scoring is based on the idea that the content assessed has a single dimension. In other words, IRT assumes that there is a strong, single construct that underlies the performance of all items. This unidimensionality represents consistency in the content assessed. A test that lacks unidimensionality may produce estimates of a student's achievement that are not as reliable as a test that assesses only a single dimension.

Quality Assurance Measures—Studies of the unidimensionality of each assessment, conducted prior to the first operational test administration for each subject area, confirm that each test fits the IRT model. For this purpose, the Q3 statistic, which is an index for how well the construct is unidimensional, is calculated across all item pairs and the results are evaluated to ensure that a single construct underlies behind what is being measured. Also, Confirmatory Factor Analysis (CFA), which is a statistical technique used to verify the factorial structure of a set of observed variables, is conducted to test the hypothesis that a relationship between observed variables and underlying single latent constructs exists.

3.1.7. Released Test Items

Beginning in the summer of 2024, teachers, parents, and other stakeholders will have the opportunity to review some of the operational test content used in Florida’s statewide assessments. [Section 1008.22\(8\)](#), F.S., requires the Department to publish each statewide, standardized assessment administered, excluding retakes, at least once on a triennial basis, with the initial publication occurring no later than June 30, 2024. The initial publication of assessments must include, at a minimum, the grade 3 ELA Reading and Mathematics assessments, the grade 10 ELA Reading assessment, and the Algebra 1 EOC assessment. Additionally, the Department plans to release B.E.S.T. Writing prompts and individual student writing responses each year.

For ELA Reading and Writing assessments, the Department will release public domain and commissioned passages and provide references to permissioned passages, which will not be able to be released as part of the interpretive product.

Each released test will include content that represents an operational test blueprint for each respective grade and subject and, per statute, released content must have appeared on tests in the administration year immediately preceding release. Because Florida’s statewide assessments are computer-adaptive, this means that items used during any administration in the previous school year may appear on the released forms.

4.0 Test Administration

It is only through a standardized and secure administration process that the statewide assessments can provide an accurate representation of student achievement. It is this standardization that makes comparisons across schools and years possible.

Quality Assurance Measures—Detailed information relating to test administration is provided in the test administration manuals. The manuals provide the requirements for test administrators who administer the test, school assessment coordinators who organize the administration in their schools, and district assessment coordinators who coordinate the assessment program for their districts.

4.1. Schedules

Districts establish testing dates within a statewide window for each administration, based on factors such as computer availability and scheduling needs. The state-established testing windows for the Statewide Assessments Program are posted on the [FDOE website](#).

4.2. Computer-Based Testing (CBT)

All grades 3–10 assessments and EOC assessments are computer-adaptive and will be administered and reported using the following systems:

Online Management/Student Data	Test Information Distribution Engine (TIDE)
Online Test Delivery	Test Delivery System (TDS) with Secure Browser
Results Reporting	Florida Reporting System (FRS)
Results for Parents	Family Portal

The contractor provides a secure, online testing platform to deliver computer-based tests to students. Students log in using a test ticket that contains their unique login information for the test they will take and navigate through and interact with the test using various tools and features. Computer-based sample items are available on the [Florida Statewide Assessments Portal](#).

For each test, a paper-based version is available as an accommodation for eligible students. Accommodations are discussed in detail later in this section.

4.3. Timing

Because the statewide assessments are not “speeded” tests (finishing in a set amount of time is not part of what is being assessed), the time allotted is designed to give students enough time to respond to all items. The chart below details the minutes per session, total minutes per test, and what extended time is available to students who have not completed the assessment in the standard session time.

Table 5: Florida Statewide Assessment Testing Times

Assessment	Grade Level(s)	Minutes Per Session
FAST ELA Reading	3–10	120*
FAST Mathematics	3–5	100*
	6–8	120*
Statewide Science	5 & 8	160*
EOC Assessments	N/A	160*
B.E.S.T. Writing	4–10	120**

*Students who are still working at the end of the allotted time for the session may continue working up to the length of a typical school day.

**Students who are still working at the end of the allotted time for the session may continue working up to half the length of a typical school day.

4.4. Personnel

State level—FDOE’s Bureau of K–12 Student Assessment employs a Test Administration team of trained staff to develop resources (such as manuals, user guides, and training materials), deliver instructions and policy information, and provide support for district-level staff during testing. State personnel also serve as liaisons between districts and the testing contractor.

District level—The district designates one of its employees as the district assessment coordinator to act as the point of contact between FDOE, the contractor, and the district’s schools. In larger districts, multiple staff members assist in supporting schools with test administration.

School level—The school designates an employee, typically a school administrator or guidance counselor, as the school assessment coordinator to act as the point of contact between the district and the school. A technology coordinator is also designated at the school level to ensure that technology setup is completed correctly and that any technology-related issues are addressed during testing.

Testing session—Test administrators (TAs) supervise test sessions. TAs must be employees of the school district and are usually classroom teachers. They must receive training prior to each test administration window and are required to remain in the testing room at all times.

Proctors are additional monitors in testing rooms and are recommended in all testing rooms, but they are only required when the number of students in the testing room exceeds 25. School personnel and non-school personnel may be trained as proctors. Prior to testing, proctors must be informed of their duties and of the appropriate test security policies and procedures. School personnel proctor duties may include preparing and distributing secure test materials. Non-school personnel may assist TAs during test administration; however, they may not participate in any of the test administration procedures (e.g., distributing and collecting secure test materials, providing accommodations). Volunteers (e.g., parents, retired teachers) may be trained as proctors and may perform non-school personnel duties.

TAs do not administer tests to their family members. Students related to their assigned TA are reassigned to an alternate TA. In addition, a student's parent/guardian may not be present in that student's testing room.

All personnel involved in the preparation or administration of statewide assessments must be adequately trained and aware of all test security policies and procedures. A *Test Administration and Security Agreement* must be signed by all testing personnel. In addition, TAs must sign a *Test Administrator Prohibited Activities Agreement*, which provides a list of actions to avoid in order to prevent student test invalidation and/or investigation for misconduct.

4.5. Test Security

Test security is an important part of maintaining the integrity of test content, test administration, and the reliability of results. Policies and procedures are in place before, during, and after testing to ensure valid test administrations, and strict processes must be followed if a breach in security is identified.

Some materials that are considered “secure” are test and response books, CBT test tickets, used work folders, used worksheets, and used planning sheets. Secure materials must be maintained and tracked by the school assessment coordinator and placed in locked storage when not in use. Paper-based secure materials, such as accommodated paper-based forms, are coded with unique security numbers for tracking purposes. A Test Materials Chain of Custody Form must be maintained at each school, listing individuals with access to the materials, as well as dates and times that the materials are checked out or returned. Secure materials must never be left unsecured and must not remain in classrooms or be taken off the school's campus overnight. No more than three people should have access to the locked storage rooms.

All content within Florida's statewide assessments is secure. Test items, which result from the detailed, rigorous development process described in previous sections, may be reused on future test forms. Because of this, TAs are prohibited from viewing test content during testing, and students are asked not to reveal test content after the administration. For more information on test security, see the links to the test security statute and rule in Appendix B.

4.6. Administration Procedures

The goal of any statewide assessment administration is to promote the validity and reliability of the test scores. While validity (ensuring that assessments are aligned to the content they are intended to measure) and reliability (ensuring that assessments consistently measure what they intend to measure) are largely achieved through the high-quality work described in the test development section, valid administration of the assessments is also a crucial component to ensure that student scores are an accurate measure of ability. **Standardization** provides that students take tests under the same conditions to the greatest extent possible so that scores are comparable. For instance, if a group of students is provided only 40 minutes for a 100-minute test session, that group of students was disadvantaged compared to students allotted the full amount of time, and the scores of the two groups cannot be compared. The policies and procedures for test administration established by FDOE are to promote best practices in standardization and reduce testing irregularities.

Each test session must be conducted according to guidelines established in the test administration manuals. TAs must ensure that the room is prepared and all materials are ready prior to each test session. Checklists are provided in the appendices of the manuals to ensure school staff, including TAs, complete all required tasks before, during, and after testing.

Tests should be administered in a room that has comfortable seating, good lighting, and an appropriate temperature. The room should be adequately ventilated and free of distractions. TAs must remove or cover all visual aids in the room, such as posters showing word lists, and arrange the room so that each student will have enough workspace for test materials. Students must not be facing each other when seated at tables, and schools should avoid seating (e.g., stadium, staggered) that allows them to easily view other students' answers.

The following additional materials are provided for certain CBT tests:

- [Planning Sheets](#) are provided for all students participating in B.E.S.T. Writing. Students may use their planning sheets to plan and prewrite their responses.
- [CBT Worksheets](#) are provided to all students taking a grade 8 Science, Civics EOC, and U.S. History EOC to take notes during the test.
- [CBT Work Folders](#) are provided to all students taking a B.E.S.T. EOC or the Biology 1 EOC to work the problems in the test.
- [Scratch Paper](#) is provided to all students taking a FAST Mathematics test.
- Online calculators ([four-function](#) and [scientific](#)) are provided for grades 6–8 Mathematics, Algebra 1 and Geometry EOCs, Biology 1 EOC, and the Grade 8 Statewide Science Assessment. Calculators are not permitted for grades 3–5 Mathematics. The [Calculator and Reference Sheet Policies](#) document describes which calculator may be used for each assessment.
- [Reference Sheets](#) are available in the online platform for grades 4–8 FAST Mathematics and B.E.S.T. Algebra 1 and Geometry EOCs.
- The [Periodic Table of the Elements](#) is available in the online platform for grade 8 Science and Biology 1 EOC.

For each assessment, test administration scripts are provided that include instructions and policy reminders for students. The scripts promote standardization by ensuring that all students receive the same information immediately prior to testing. Once students are in a testing room, the TA will read the appropriate script verbatim to students. Each script includes the following information:

- **Electronic Devices Policy**—To ensure test security, FDOE employs a strict “no electronic devices” policy during testing. A detailed list of prohibited devices is read to all students and they are reminded that being found in violation of the policy is cause for test invalidation. Then, students are instructed to raise their hands if they have an electronic device with them. The TA will then follow his or her district’s procedure for devices (e.g., collect them until after testing).
- **CBT Instructions**—directions for logging into a computer-based test, accessing the help menu in the test, and navigating, pausing, and submitting the test.
- **PBT Instructions**—directions for completing the front cover and unsealing the test documents.
- **Testing Rules and Testing Rules Acknowledgment**—Students are read the following list prior to testing (this example is from Writing; other assessments have minor differences):

Remember:

- *You may not have a cell phone during testing.*
- *Do not talk to other students or make any disturbance.*
- *Do not look at another student’s test materials.*
- *Do not ask for or provide help in answering any test questions.*
- *Use only approved materials for taking notes.*
- *Do not have access to or use any electronic or recording devices at any time during this test session, including during breaks.*

Because the content in all statewide assessments is secure, you may not discuss or reveal details about test items or content after the test. This includes any type of electronic communication, such as texting, emailing, or posting to social media sites.

Once this portion of the script is read aloud, students are asked to select a check box to indicate that they have read and understood the Testing Rules Acknowledgment that states, *I understand these testing rules. If I do not follow these rules, my test score may be invalidated.* (Students participating in paper-based accommodated testing sign below the Testing Rules Acknowledgment.)

- **Timing**—Test administration scripts provide the amount of time that must be allotted for each session and scheduling guidance for breaks and, if applicable, additional time.

Scripts are available in each test administration manual and are provided as a separate resource for administrations with certain accommodations. These resources are available on the [Florida Statewide Assessments Portal](#).

4.7. Accommodations

Many students with disabilities require testing accommodations in order to participate in testing programs on an equal basis with their peers who possess no disability. Accommodations provide students with the ability to demonstrate what they know and can do without the limitations of a disability.

The [Individuals with Disabilities Education Act \(IDEA\)](#) states that students with disabilities who have current Individual Education Plans (IEPs) or Section 504 Plans may receive assessment accommodations unique to their specific needs, including, flexible presentation, flexible responding, flexible scheduling, flexible setting, and assistive devices and tools. In order to be approved for use during testing, a unique accommodation must be documented on the IEP or Section 504 Plan, must be used regularly by the student in the classroom, and must not negate the validity or threaten the security of the assessment.

Students who are classified as English Language Learners (ELLs) must also participate in the statewide assessments program as required by section 1008.22, F.S. ELLs may receive approved accommodations for assessments that include flexible scheduling, flexible setting, assistance in heritage language, and an approved dictionary/glossary.

[The Statewide Assessments Accommodations Guide](#) lists allowable accommodations and related policies. The guide provides comprehensive information to support successful test administrations to students with disabilities and ELLs. If TAs are administering tests to ELLs, recently exited ELLs, or students with disabilities who have allowable accommodations documented on an IEP or Section 504 Plan, care must be taken to provide the exact accommodation needed.

4.8. Test Invalidation

Certain policy violations may result in test invalidation. If a test irregularity occurs that may cause the student's results to not be a valid representation of his or her ability, the test is not scored and no results are reported. Some common reasons a student's test may be invalidated are:

- Possession of an electronic device during testing,
- Not working independently,
- Taking the wrong grade level assessment,
- Being provided with accommodations that are not allowable or are not listed on the student's IEP or Section 504 Plan,
- Not being provided an accommodation listed on the student's IEP or Section 504 Plan, or
- Being given unauthorized assistance during testing.

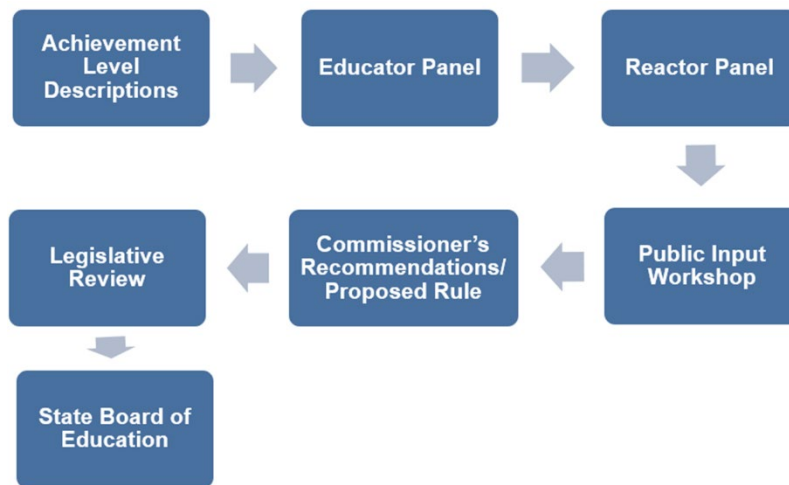
5.0 Scoring the Assessments

The scoring process should be viewed in terms of item scoring and whole-test scoring. This distinction is necessary because the discussion of item scoring focuses on the methods used to rate student responses to individual items, whereas the discussion of whole-test scoring focuses on the statistical methods used to derive scale scores for the test overall.

This chapter is divided into two sections, one dealing with the process and methods for scoring items and the other describing the methods used to generate scores for the test as a whole, including scale scores and Achievement Level classifications.

5.1. Standard Setting

As assessments are implemented for the first time, a standard setting process is required. FDOE seeks input from educators, school districts, and business and community leaders to determine the proposed Achievement Level standards for the new statewide assessments, and the State Board of Education ultimately reviews the recommendations and establishes the standards in State Board of Education Rule.



As shown in the graphic below, the first stage of this process involves writing [Achievement Level Descriptions](#) (ALDs). ALDs describe a student’s level of achievement (e.g., Below Grade Level, On Grade Level, Proficient) on a large-scale assessment. FDOE convenes panels of Florida educators to develop ALDs to guide participants during the standard-setting process for its statewide assessments, offer score interpretation on student reports, and further teacher understanding of expectations for the progressions of student performance at each achievement level. The purpose of the ALD development framework is to enable valid inferences about student content area knowledge and skill in relation to a state’s content standards measured on a statewide assessment.

Next, educator panels are again convened to propose recommended cut scores using the ALDs. These panels are comprised of individuals with content expertise who understand expectations

for Florida students and focus on content when making recommendations. Their recommendations are then presented to a reactor panel composed of community leaders who represent various business, community, and educational groups who consider policy and the impact of the proposed scores when making their recommendations to the Commissioner of Education.

Finally, the Commissioner considers the recommendations of the educator panel, the reactor panel, and the public before making final recommendations to the State Board of Education for adoption of the new standards.

In October 2023, the State Board of Education adopted new cut scores on the B.E.S.T. scale for the FAST and B.E.S.T. assessments, and these new scores took effect for assessments taken in winter 2023 and beyond. For more information about the implementation of the B.E.S.T. score scale, please see the [FAQ](#) posted on the FDOE website.

Standard setting was conducted for the Statewide Science, Biology 1 EOC, and U.S. History EOC assessments in 2013, and the Civics EOC assessment in 2014. The [Standard Setting page](#) on FDOE’s website provides standard setting information and resources for Florida’s statewide assessments in chronological order, beginning with the most recently adopted assessments.

5.2. Scoring Different Item Types

5.2.1. Machine/Electronic Scoring

Most item types, such as multiple-choice, multiselect, and matching, have one correct answer or answer set. For most two-part items, students must answer both parts correctly to receive full credit for that item. In mathematics assessments, some two-part items are worth two points and students can receive partial credit if they only answer one part correctly.

Equation response items are scored using rubrics that contain all possible formats of a correct response (e.g., “ $y = 2 + x$ ” and “ $2 + x = y$ ”) to ensure students receive credit. The rubrics are reviewed by content specialists and Florida educators on Item Review Committees, then again after field testing by Rubric Validation committees to ensure that all possible correct answers are included in the rubrics.

Numerous checks are incorporated in the scoring program to alert scoring personnel to any possible problems with an item, such as when a large number of otherwise high-achieving students choose or provide an answer that was not originally identified as correct. These situations lead scoring personnel to investigate whether there is a misleading distractor or more than one correct answer to a multiple-choice item or whether the list of acceptable answers to equation editor items may need to be expanded.

Quality Assurance Measures: Statistical Reviews—Field-test items are embedded into each spring’s tests and randomly administered to a large representative group of students, reflecting characteristics of the student population in the state. The resulting data set is used by Rubric Validation committees to determine the final rubric for each item. After this step, field-test items are calibrated and placed onto the item bank scale. Psychometricians initially review their statistics, such as p -values, biserial and poly-serial correlations, IRT statistics, item-fit, and DIF indices (but not the content of items) and decide whether items should be kept in the bank or removed. In the next step, items are taken to the Data Review Committee meeting to consider both their content and statistical features. Items which successfully go through all of these steps are moved into the operational item pool and become candidates to be used as an operational item in future test administrations.

5.2.2. Hybrid Scoring – B.E.S.T. Writing

Scoring of B.E.S.T. Writing involves the use of a large representative sample of Florida student responses scored by highly trained and qualified human scorers to train an automated scoring (AS) engine, which then scores the majority of student responses. This model, called “hybrid scoring,” has been successfully used in Florida previously, and is widely used across the country in a number of statewide assessment programs. Hybrid scoring allows for a faster scoring process while maintaining Florida’s high validity criteria.

All human scorers are trained by FDOE ELA content specialists and the Department’s scoring sub-contractor, Data Recognition Corporation (DRC). The AS engine is trained using Florida student responses that were previously scored by at least two humans after a field-test administration. Expert Florida educators review and approve training materials used in both the stand-alone field-test scoring process and the operational B.E.S.T. Writing assessment given each spring. Once trained, the AS engine produces a response score as well as a “confidence” score indicating the level of certainty that the score is accurate. In addition to quality control checks on a random sample of all responses scored by the AS engine, up to 40% of all student responses are scored by humans. These are responses that are identified as unusual or that do not fully meet the Department’s confidence criteria for accepting the AS engine score as the score of record.

After scores are reported to districts – along with individual student responses, prompts, source material, and anchor sets with annotations linking anchor responses to rubric scores – districts have the opportunity to submit score inquiries to verify scores for responses that district assessment and instructional staff deem to warrant further review.

In summary, Florida relies on trained expert human scorers who are the foundation of B.E.S.T. Writing scoring, and these experts are involved throughout the entire scoring and oversight process from beginning to end, including the automated scoring process.

To guide human scorers in rating responses, they have the following tools and references at their disposal:

- A general scoring rubric with descriptions of work demonstrative of each point on the scale in each of the three following domains:
 - Purpose/Structure
 - Development
 - Language
- Anchor papers with annotations—Actual, unedited student responses that illustrate typical performance for each point in each domain. Each student response is annotated with a rationale for the score given. Anchor papers are also called rangefinder papers.

The anchor papers are developed initially by Florida educators serving on Rangefinder Committees and then reviewed and refined by FDOE and the scoring contractor on Rangefinder Review Committees. After Writing prompts are selected for use as operational items, Rangefinder Review Committees review the scoring guides and training materials originally established by the Rangefinder Committees. Each Rangefinder Committee is comprised of Florida educators, including teachers, school and district curriculum specialists, and university faculty.

Quality Assurance Measures for Handscoring—Numerous measures are in place to ensure scoring accuracy and consistency.

- **Backreading**—Team leaders (and scoring directors, as needed) check the work of individual scorers to ensure that they are scoring responses in accordance with the established guidelines. Team leaders, scoring directors, and assistant scoring directors read behind all scorers throughout the scoring session. This is called backreading, and it is done with more frequency at the beginning of the scoring session to identify scorers who may need additional training and monitoring. Team leaders, scoring directors, and assistant scoring directors ask scorers to review responses that were scored incorrectly and then provide guidance on how to score more accurately.
- **Daily Review of Training Materials**—At the beginning of each scoring session, team members spend at least 15 minutes reviewing their training materials and scoring guidelines, including anchor papers and item-specific criteria.
- **Calibration Sessions**—Scorers in teams working on the same Writing prompt meet in a full group daily for the purpose of calibration. Calibration responses are identified by team leaders and scoring directors and approved by FDOE. Each calibration set will include one to three student responses and are intended to target trends or issues within the room. Scoring directors discuss calibration sets with the scorers after each calibration set is taken.

- **Retraining**—Retraining is conducted for scorers whose scores are identified as inaccurate or who fall below acceptable standards. Papers used for retraining are identified by the team leaders and scoring directors based on the scorers' validity accuracy (validity outlined below) and may include validity papers, anchor review, and rubric review. If retraining is unsuccessful, scorers are dismissed from the program, and responses scored by that individual are reset to be scored again.
- **Validity and Reliability Reports**—Embedded in the flow of student responses that scorers score at their work stations are responses for which scores have already been established by the Rangefinder and Rangefinder Review Committees, as well as FDOE content specialists. Comparisons of the scores assigned by a scorer with the established scores are compiled as validity reports and presented to scoring directors and FDOE staff throughout the scoring sessions. From the validity reports, scoring directors can see which responses are most often scored incorrectly and which scorers are most often in disagreement with the established scores. Reliability (consistency) of handscoring is monitored using reports of inter-rater reliability. Each scorer's (or rater's) score on a student response is compared to the other score given to that response. A cumulative percent of agreement between the two scores on every response (as opposed to validity responses only) is reported for each scorer as the inter-rater reliability percent. The information on this report indicates whether a scorer is agreeing with other scorers scoring the same responses. Analysis of the report is used to determine if a scorer or group of scorers is drifting from the established guidelines and requires additional training.

Quality Assurance Measures for Automated Scoring

- **Confidence Scores**—When the scoring engine determines a score for a student response, it also issues a confidence score that asserts how confident the engine is in the score it provided. For instance, if a student provides an unusual response that the engine is unsure how to score, it will generate a low confidence score. Responses with low confidence scores are always rescored by a human reader to ensure that the score assigned to the student is correct.
- **Validity**—A random 5% of papers at each grade level are routed to human scorers for the purpose of tracking engine and human agreement. If disagreement between the engine scores and humans scores reveal that the engine is assigning inaccurate scores to responses, the engine can be retrained using training papers from the original training set and additional papers scored by humans in the current operational administration. If this occurs, papers scored by the engine up to that point can be reset to be scored again.
- **Quadratic Weighted Kappa (QWK)**—A statistical measure that shows whether the engine human agreement rate is greater than chance. In other words, this statistic is able to show whether the human scorers and the engine are agreeing due to accuracy of their scoring or due to chance.

- **Standardized Mean Difference (SMD)**—During scoring, the means and standard deviations of human scorers and the engine are monitored for differences.

5.2.3. Reported Measures

Statewide ELA Reading, Mathematics, Science and Social Studies assessments are reported with the following information:

- **Scale scores**—Each assessment has a score scale. Student performance is represented by where a student’s score falls on the scale.
- **Achievement Levels**—Scale scores fall into five achievement levels that provide descriptions of a student’s performance.
- **Achievement levels by reporting category**—The student’s achievement level (below, at/near, or above the standard) for each reporting category on the test is reported. These classifications indicate a student’s level of success with items that assess the benchmarks within each category.

The lowest scale score in Achievement Level 3 is considered the “passing” score for each assessment; however, the only state-level requirements apply to the Grade 3 ELA Reading, Grade 10 ELA Reading, and Algebra 1 EOC Assessments. Students in grade 3 must achieve a level 2 or higher on the ELA assessment to be promoted to grade 4. Passing the Grade 10 ELA and Algebra 1 assessments is required for graduation with a standard high school diploma. Students who participated in the first administration of one of these graduation assessments, as well as students with older graduation requirements, are eligible to use an “alternate passing score” linked to the passing score on the previous assessment required for graduation.

For more information about graduation requirements, including comparative and concordant scores on other assessments that may be used to satisfy the grade 10 ELA and Algebra 1 EOC assessment requirements, please see [Graduation Requirements for Florida’s Statewide Assessments](#).

For B.E.S.T. Writing, the following information is reported:

- **Scale scores**—Each assessment has a score scale. Scale scores indicate where a student’s score falls on the scale. When a scale score is evaluated together with the grade-level cut score, the scale score indicates whether a student’s writing performance is on or below grade level.
- **Raw scores**—These scores reflect how many points a student earned out of 12 possible points.
- **Domain scores**—Each rubric is broken down into three domains that measure different skills. These scores reflect how many points (out of four possible points) students earned in each domain.

5.2.4. IRT Scoring

Scale scores are the result of a statistical optimization process where student responses and item parameter estimates are jointly evaluated to derive the most likely score, comparable across test years. As described in the IRT Framework section on page 23 the IRT model assumes that student responses to items are the result of underlying levels of knowledge and skills. Also, item parameter estimates are independent from the population that they are derived from. When these two pieces of information are assessed together via a maximum likelihood algorithm through an optimization, unbiased estimates of student abilities or scale scores are generated. The goal of the assessment program and of the quality control process described in this guide is to accurately report a score as close to the true level of ability as possible.

Another key feature of the IRT model is that ability and item difficulty exist on a single dimension so that students with low abilities will generally succeed on less difficult items, students with moderate abilities will typically succeed on items with low to moderate difficulty, and students with high abilities will succeed on items at all levels of difficulty. Ideally, any test constructed using the IRT model will include items that clearly distinguish between students with increasing levels of ability.

Two important aspects of IRT processing contrast with traditional methods of test scoring. One aspect is that items are given different considerations based on their differing IRT parameter estimates when calculating the overall score. For example, relatively more consideration might be given to items with a greater discrimination (a high a -parameter estimate) and relatively less consideration might be given to items on which a lot of pseudo-guessing occurs (a high c -parameter estimate). In situations like these, different considerations apply in the same way to the calculation of scores for all students.

Another important contrast between IRT scoring and traditional methods is the use of *pattern scoring*. That is, the pattern of answers provided by a student is analyzed in combination with the IRT item parameter estimates. In other words, information about the pattern of answers (which questions were answered) and the statistical qualities of test items (discrimination, guessing, and difficulty) are evaluated together to determine the scoring weights for each item and the likelihood of individual student score. As a result of this method of scoring, students with the same raw score may have similar, but not necessarily identical, scale scores. Different scale scores result because each item contributes uniquely to a student's overall scale score. Students who correctly answer exactly the same items would, of course, receive the same scale score.

IRT pattern scoring is used for statewide assessments because it produces more accurate depictions of students' true levels of ability (knowledge and skill). Using IRT pattern scoring is an important method of ensuring the most accurate measure of student achievement possible.

Appendix A of this document provides additional technical information about the statistical indicators used in the scoring process.

Achievement Level Classifications

Based on students' scale scores, they are assigned into one of five Achievement Level classifications. Achievement Levels are ranges of scores within each assessment's scale. The cut

point scores (numerical borders) between each level were established as part of the standard setting process (see Section 5.1, Standard Setting). The levels range from the lowest level (Level 1) to the highest level (Level 5).

Determining a student's Achievement Level classification involves locating the score in one of the five Achievement Levels. Achievement Level classifications provide a clearer statement than the scale score in regard to a student's performance. For schools, districts, and the state, monitoring changes in the percentages of students in each level provides a convenient method of comparing progress over time.

Quality Assurance Measures—One statistical review conducted after operational testing evaluates the accuracy and consistency of the Achievement Level classifications. Because placement in a specified Achievement Level is a requirement for high school graduation (on grade 10 ELA and Algebra 1 EOC) and is also used in decisions regarding promotion from grade 3 to grade 4, the accuracy and consistency of these classifications is extremely important.

6.0 Reporting Results

For each test administration, reports containing assessment results are sent to four major audiences: students and their parents/guardians, school administrators, district administrators, and state-level administrators and policymakers. Each spring, FDOE also makes results available to the general public on the FDOE website.

6.1. Reports

Florida Reporting System (FRS)

FRS is a dynamic platform developed and provided by the Department's assessment contractor, Cambium Assessment, Inc., where school-, district-, and state-level users can access student results for statewide assessments. Because the tests are computer-adaptive, scores are reported in FRS within 24 hours after students complete testing (with the exception of B.E.S.T. Writing). Teachers with a TA role can view data for all students in their rosters who have completed assessments. School-level users (Private School Administrators, School Assessment Coordinators, School Administrators, and School Reporting Access) can view data for all students in their schools who have completed assessments. District-level users (District Assessment Coordinators, District Administrators, and District Reporting Access) can view data for all students in their district who have completed assessments.

For a comprehensive look at what information is available in FRS, please see [Understanding FAST and B.E.S.T. Reports for Teachers](#).

Family Portal

Families can access their student's results in the Family Portal using the login information provided by the student's school, which includes the student's unique six-digit access code. Families can access the portal directly from the FAST portal or through their district's Student Information System (SIS). In the Family Portal dashboard, users can see and print their student's scale score and achievement level, a chart indicating the student's scale score and where it falls in the achievement level, and previous assessment results, if available.

For a comprehensive look at what information is provided for parents and families, please see [Understanding FAST and B.E.S.T. Reports for Families](#).

Individual Student Reports

The **Individual Student Report** is a dynamically-generated report in FRS and the Family Portal that provides general information about each student's results, including the scale score, achievement level, achievement level by reporting category, and previous performance (for ELA Reading and Mathematics only, if available). The report also indicates how the student's performance compares to that of other students who took the same test in the same school, district, and the state, and provides helpful resources.

For FAST ELA and Mathematics and B.E.S.T. EOC assessments, the Individual Student Report contains additional language that indicates whether a student performed below, at/near, or above the standard and suggests ways that parents can help students make progress in their learning. This enhanced language is being developed for Science and Social Studies reports for the 2024–25 school year.

Quality Assurance Measures: Score Inquiry System—If a score for a student is missing, not reported, or is thought to be anomalous, the student’s school district may submit a score inquiry to FDOE. FDOE and the contractor will research the case, provide the school district with the results, and update records in late reporting, if necessary. Examples of inquiries are a missing score or confirming with ELA specialists that a student writing response received the correct score.

6.2. Know Your Data: Advanced Reports

Florida’s [Know Your Data Portal](#) connects people with reports and statistical information about statewide assessment results, which are available for the most recent five years of an assessment. The portal allows users to generate reports that offer overall and demographic information in a variety of formats including tables, maps, achievement gap graphs, and custom reports based on a user’s interests in education-related data. The goal of the portal is to increase the advancement of Florida students and schools by providing access to information that will facilitate important decisions about education in Florida. Users are encouraged to review the [User Manual](#) to understand the functionality and full capabilities of the information portal. In addition to assessment results, Know Your Data provides advanced users with access to other education information and outcomes, such as enrollment data and high school graduation rates.

6.3. Know Your Schools: Florida School, District, and State Report Cards

School, District, and State Report Cards were developed by the Florida Department of Education to meet the requirements set forth by the *Elementary and Secondary Education Act of 1965* as amended by the *Every Student Succeeds Act* (ESSA) and are available in the [Know Your Schools Portal](#), also referred to as EduData. These report cards serve as a valuable resource for parents, educators, and stakeholders by offering greater transparency into how Florida’s districts and schools are doing with regard to student achievement and success. The Report Cards provide performance data but also contain several other features including a State Report Card, educator qualifications and equity, per-pupil expenditures, and national data. The department will continue to update information in the report cards as new data become available, as well as develop new enhancements based on user feedback received through the portal. News and updates about ongoing enhancements will be provided to those who subscribe through the portal.

7.0 Glossary

Terms in boldface type appear within the glossary as separate entries.

Achievement Levels: Five categories of achievement that represent the success students demonstrate with content assessed on the **FAST**, **B.E.S.T.**, and/or **Science/Social Studies** assessments. Achievement Levels are established using the input of classroom teachers, curriculum specialists, education administrators, and other interested citizens. These professionals helped FDOE identify the score ranges for each Achievement Level. The Achievement Levels are helpful in interpreting what a student's **scale score** represents.

Backreading: Process in which scoring supervisors check the work of individual scorers to ensure that they are scoring responses in accordance with the established guidelines.

Benchmarks for Excellent Student Thinking (B.E.S.T.) Assessments: Assessments include grades 4 through 10 Writing and end-of-course (EOC) assessments in Algebra 1 and Geometry.

Bias: Advantage or disadvantage conferred upon groups of students because of certain personal characteristics (such as sex, ethnicity, disability status, or English language learner status) unrelated to an understanding of the content.

Bias Review Committee: Committee composed of educators from Florida school districts and universities who look for any items, prompts, graphics, or passages that might provide an advantage or disadvantage to a student with certain personal characteristics.

Calibration Sessions: Sessions in which handscorers meet as a team to review scoring guidelines.

Community Sensitivity Committee: Committee made up of Florida citizens associated with a variety of organizations and institutions who are asked to consider whether the subject matter and language of test items, writing prompts, graphics, or reading passages will be acceptable to students, their parents, and other members of Florida communities.

Content Advisory Committee: A group composed of 15–24 subject area specialists from schools, districts, and universities across Florida who periodically revise specifications to provide new **sample items** and **reading passages**.

Content Area: The information or skills contained in an area of study. The content areas (or subject areas) assessed are writing, reading, mathematics, science, and social studies.

Criterion-Referenced Test (CRT): An assessment where an individual's performance is compared to a specific learning objective or performance standard and not to the performance of other students. Criterion-referenced tests show how well students performed on specific goals or **standards** rather than just telling how their performance compares to a norm group of students nationally or locally.

Cut Point Scores: Scale scores that mark the boundaries between different **achievement levels**.

Differential Item Functioning: Statistical property of an item that indicates how likely it is for students from distinct subgroups, such as sex, ethnicity, etc., possessing similar skills, responds to an item differently. It occurs when individuals from different subgroups, with comparable skill levels, do not have an equal likelihood of answering an item correctly.

Drag-and-Drop Hot Text: Items with a bank of words, phrases, objects, or sentences that may be dragged into a response area.

Equating: A statistical process generating test scores that are interchangeable across different test forms and years.

Equation Editor: Available in Mathematics only, an item in which a student is presented with a toolbar that includes a variety of mathematical symbols that can be used to create a response.

Expert Review Committee: Panel of university-level and practicing research scientists and legal experts who ensure the content accuracy of the test items in fields where information can change over time.

Evidence-Based Selected Response: Two-part item in which a student is directed to select the correct answers from Part A and Part B.

Exceptional Student Education (ESE): Special educational services that are provided to eligible students (e.g., visually impaired, hearing impaired). These services are required by federal law and provided to Florida students according to the State Board of Education Rule 6A-6.0331, Florida Administrative Code (F.A.C.).

Field-Test Item: Items included in an assessment for **item** development purposes only. Student response data are reviewed to determine whether a field-test item would be a useful **operational item**. Field-test items do not count toward student scores.

Florida Assessment of Student Thinking (FAST) Assessments: The FAST, which includes VPK through grade 10 Reading and VPK through grade 8 Mathematics assessments, are administered as progress monitoring assessments three times per year.

Florida Statewide Assessments Portal: Resources and information for students, parents, members of the public, and district and school personnel are located on the Florida Statewide Assessments Portal, which is accessed at www.FSAssessments.org.

Hot Text: See **Selectable Hot Text** or **Drag-and-Drop Hot Text**.

Individual Education Plan (IEP): Describes special education services provided as part of **Exceptional Student Education**. Also specifies the testing accommodations a student needs for classroom instruction and assessments.

Item: Any test question or task for which a separate score is awarded.

Item Bank: Database of **field-test** or **operational items**.

Item Content Review Committee: Committee whose members determine whether the passages, graphics, and items are appropriate for the proposed grade levels.

Item Response Theory (IRT): A statistical model describing the relationship between for student's ability level and response to a test **item**. Based on the idea that the likelihood of student success on an **item** is the result of the student's true level of ability and three characteristics of the **item**: ability of the **item** to differentiate between students at different **Achievement Levels** (the a -parameter), difficulty of the **item** (the b -parameter), and the effectiveness of pseudo-guessing (the c -parameter, for **multiple-choice items** only). Used in **item** and test development and as the basis of generating **scale scores**.

Multimedia: Available in ELA only, an **item** in which technology-enhanced content may include multimedia elements such as audio clips, slideshows, or animations.

Multiselect: **Item** in which a student is directed to select a specific number of or all of the correct answers from among the options provided.

Multiple-Choice Item: **Items** that present students with several options from which to choose one response.

Open Response: **Item** in which a student uses the keyboard to enter a response into a text field.

Operational Item: An **item** that counts toward a student's score.

Pattern Scoring: A statistical optimization process where student responses and item parameter estimates are jointly evaluated to derive the most likely scale score

Performance Tasks: **Items** that require students to provide either a short or extended written response.

Prompt: The topic a student is given to provide a response for **B.E.S.T.** Writing.

Psychometrics: The field of study generally covering specialized fields within psychology and education devoted to testing, assessment, measurement, and related activities.

Rangefinder: Student responses to **prompts** or **performance tasks** used to illustrate score points. Rangefinding is the processing of identifying these student responses.

Rangefinder Review Committee: Members examine a representative set of student responses from Writing prompt field tests to establish scoring guidelines.

Reading Passages: There are two types of reading passages: informational and literary.

Reliability: Desired characteristic of a test. Achieved when measurement error is minimized and the test score is close to the **true score**.

Retake: Alternate assessment given to those who do not achieve the passing score required for high school graduation. Retake tests are available for **FAST** Grade 10 ELA Reading and **B.E.S.T.** Algebra 1 EOC.

Rubric Validation Committee: Members review all responses to field-tested equation editor items to determine whether all possible correct answers have been included in the scoring key.

Sample Items: Sample items are available for students and the general public to preview item types and functionality in the online platform.

Scale Score: Score used to report student results for the entire test. The scale score is the result of **IRT** processing and **equating**.

Science/Social Studies Assessments: Students in grades 5 and 8 participate in the statewide science assessments. End-of-course assessments include Biology 1, Civics, and U.S. History.

Section 504: Special classification of students as defined in Section 504 of the Rehabilitation Act of 1973. Testing accommodations are permitted for students who meet the Section 504 criteria.

Secure Materials: Materials used for testing that must be kept secure and tracked using a chain of custody form. Examples include writing passage booklets used for computer-based testing and **test and response books** for paper-based accommodated tests .

Selectable Hot Text: **Item** type with highlighted words, phrases, or sentences in which a student can click to select an answer. These **items** may have one or two parts.

Standard Error of Measurement: A whole-test **reliability** indicator that is calculated using data from the entire tested population. For example, if a student were to take the same test over and over (without additional learning between the tests or without remembering any of the questions from the previous tests), the difference in the resulting test scores is called the standard error of measurement.

Standard Setting: Process in which the Department seeks input from educators, school districts, and business and community leaders to determine the proposed **Achievement Level** standards for the new statewide assessments.

Table Match: This **item** type presents options in columns and rows. Options may include numerical values, words, phrases, sentences, quotations, line/paragraph/passage numbers, or images.

Technical Advisory Committee: Committee composed of 10–15 professionals with expertise in psychometrics and/or assessment who assist the Department by reviewing technical decisions and documents, and by providing advice regarding the approaches the Department should use to analyze and report assessment data.

Technology-Enhanced Item: Technology-enhanced **items** are computer-delivered **items** that require students to interact with test content to select, construct, and/or support their answers.

Test and Response Books: For paper-based accommodated tests, students are provided test and response books that contain all of the **items**. Students respond directly in their books.

Testing Irregularities: Possible examples of testing irregularities include testing that is interrupted for an extended period of time due to a local technical malfunction or severe weather, a student copying another student's answers, or a test administrator changing students' answers. If an irregularity is found in the data, flagged student records are put on hold and FDOE staff review the data.

Universal Design: The application of these principles helps develop assessments that are usable to the greatest number of test takers, including students with disabilities and nonnative speakers of English. Universal design principles also inform decisions about test layout and design, including, but not limited to, type size, line length, spacing, and graphics.

Validity: Desired characteristic of a test. Achieved when the test actually measures what it is intended to measure.

Validity and Reliability Reports: Comparisons of the scores assigned by a scorer with the established scores are compiled as validity reports and presented to scoring directors and Department staff throughout the scoring sessions.

8.0 Guide to Useful Resources

8.1. Writing Resources

Several interpretive resources are available to provide examples and information on how Writing is scored.

- [Scoring Samplers](#) provide examples of student responses represent various combinations of the score points across the scoring domains. As a basis for developing a common understanding of the scoring criteria, an annotation follows the response to explain the prominent characteristics of the response described in the rubric.
- [Scoring Rubrics](#) describe how points are given in each domain for each mode (opinion, informative, explanatory, argumentation).

8.2. Fact Sheets

Fact sheets give a brief overview of each assessment, and are updated annually.

- [2023–2024 B.E.S.T. Algebra 1 and Geometry End-of-Course Assessments Fact Sheet](#)
- [2023–2024 B.E.S.T. Writing Fact Sheet](#)
- [2023–2024 Statewide Science Assessments Fact Sheet](#)
- [2023–2024 Science and Social Studies EOC Assessments Fact Sheet](#)
- [2023–2024 Florida Civic Literacy Examination Fact Sheet](#)

8.3. Graduation Requirements

Florida's public high school graduation requirements are specified in the following sections of Florida Statute (F.S.):

- [Section 1003.4282](#), F.S., Requirements for a standard high school diploma (effective July 1, 2013).
- [Section 1002.3105](#), F.S., Academically Challenging Curriculum to Enhance Learning (ACCEL) options.

[Graduation Requirements for Florida's Statewide Assessments](#) describes assessment requirements for graduation, including concordant and comparative score information.

8.4. Sample Test Materials

The purpose of the sample test materials is for students to become familiar with the CBT system, functionality, and item types for CBT, and the layout, format, and item types for PBT. The sample test materials are not intended to guide classroom instruction. Sample test items, answer keys, and optional scripts are available on the [portal](#). There is also a [user guide](#) for CBT sample test materials.

8.5. Technical Reports

The statewide assessments annual [technical reports](#) document methods used in test construction, psychometric properties of the tests, summaries of student results, and evidence and support for the intended uses and interpretations of assessment results.

8.6. Test Administration Manuals

Test administration manuals contain policy and procedural information for administering statewide assessments. They include instructions for school and district staff responsible for preparing for, administering, monitoring, and returning tests.

- [Test Administration Manuals](#)

In addition to the manuals, an [Accommodations Guide](#) is provided each year to offer detailed information to aid school and district staff in administering assessments to students with disabilities or English language learners who are eligible for accommodations.

8.7. Test Design Summaries and Blueprints

Test Design Summaries and Blueprints provide a map or blueprint for how each assessment is designed. The summaries show the standards assessed within each reporting category and the representation of each category, in percentages, on the test.

- [FAST Mathematics and B.E.S.T. EOCs Test Design Summaries and Blueprints](#)
- [FAST English Language Arts and B.E.S.T. Writing Test Design Summaries and Blueprints](#)
- [Science Test Design Summary and Blueprint](#)
- [Social Studies Test Design Summary and Blueprint](#)

8.8. Test Item Specifications

Test Item Specifications are based upon Florida's standards and the Florida Course Descriptions as provided in CPALMs. The Specifications are a resource that defines the content and format of the test and test items for item writers and reviewers. Each grade-level and course Specifications document indicates the alignment of items with the appropriate standards. It also serves to provide all stakeholders with information about the scope and function of the assessment program. Access to FAST and B.E.S.T. Test Item Specifications are granted to one district contact upon request to Assessment@fldoe.org, and inquiries within the district regarding the specifications are directed to that person. [Science and social studies test item specifications](#) are posted on the FDOE website.

8.9. Schedules

[Statewide Assessment Schedules](#) are established according to section 1008.22, F.S. These schedules also contain testing dates and windows for other statewide assessments, such as FAA, ACCESS for ELLs, and NAEP.

8.10. Standard Setting

As assessments are implemented for the first time, a standard setting process is required. FDOE seeks input from educators, school districts, and business and community leaders to determine the proposed Achievement Level standards for the new statewide assessments. Then, the department seeks feedback from the Legislature during a required 90-day review process, and the public is given an opportunity to submit feedback as well. Ultimately, the State Board of Education establishes the Achievement Level standards in State Board of Education Rule based on the Commissioner’s final recommendations, which take all input into consideration.

The [Standard Setting page](#) on the FDOE website provides information for each assessment in chronological order.

8.11. Understanding Reports

These documents are useful in describing the various elements of the score reports provided for students and parents, educators, and district staff.

- [Understanding FAST and B.E.S.T. Reports for Teachers and Administrators](#)
- [Understanding FAST and B.E.S.T. Reports for Families](#)

8.12. Useful Links

- The [Florida Statewide Assessments Portal](#) is the website containing resources, links, and information for FAST, B.E.S.T., Science, Social Studies, and FSA Retake assessments.
- [CPALMS](#) is the portal to Florida’s educational standards, benchmarks, and course descriptions. This website also contains links to other educational and training resources.
- The [Florida Alternate Assessment \(FAA\)](#) page provides information about the performance task and datafolio assessments administered to students with significant cognitive disabilities.
- The [ACCESS for ELLs](#) suite of assessments is administered to eligible ELL students in grades K–12. Additional information regarding Florida’s administration of ACCESS for ELLs assessments may be found on [Florida’s WIDA page](#).
- The [National Assessment of Educational Progress \(NAEP\)](#), also known as the “Nation’s Report Card,” is a congressionally mandated project administered by the National Center for Education Statistics (NCES) and overseen by the National Assessment Governing Board (NAGB). In the state of Florida, participation in NAEP is mandated by [Section 1008.22](#), F.S.

Appendix A: Statistical Indicators Used in Data Analysis

Field-test items are embedded into tests and randomly administered to a large representative group of students, reflecting characteristics of the student population in the state. The resulting data set is used by rubric validation committees to determine the final rubric for each item. After this step, field-test items are calibrated and placed onto the item bank scale. Psychometricians initially review their statistics, such as p -values, biserial and poly-serial correlations, IRT statistics, item-fit, and DIF indices (but not the content of items though) and decide whether items should be kept in the bank or removed. In the next step, items are taken to the Data Review Committee meeting to identify if there is any issue with them, considering both their content and statistical features. Items which successfully went through all of these steps are moved into operational item pool and become candidates to be used as an operational item in the future test administrations. Additional measures are generated after test administration and scoring to verify the reliability of the test and the accuracy and consistency of the Achievement Level classifications.

It is important to remember that items not meeting the criteria in any of the steps above may be rejected for use as operational items or excluded from calculation of student scores. These instances are rare because the processes of item development are carefully guided and include many quality control measures.

The following information on the various indicators is more detailed than that presented in the body of this publication. For even more detailed information, including selected data for a given year, refer to the FAST Technical Reports.

Indicator Definitions

- Differential Item Functioning (DIF)**—Indicates differences in scores between subgroups of students that are unique to the item and cannot be explained by differences between the subgroups in overall achievement. DIF statistics are calculated for sex (male versus female) and ethnic groups (e.g., White versus Hispanic), disability status (students with disabilities versus others), and English language status (English language learners versus others). Test developers typically use two types of measures of DIF, the Mantel-Haenszel chi-square statistic (Holland & Thayer, 1988)³ for single-point items and the Generalized Mantel-Haenszel chi-square statistic (Somes, 1986)⁴ for multi-point items, such as performance task items. To derive both types of measures, all students are divided into groups with similar total test scores. Within these groups, scores for each individual item are compared between subgroups of students, such as males and females, racial/ethnic groups (i.e., African American, Caucasian, and Latin American), disability

³ 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: Lawrence Erlbaum.

⁴ Somes, G. W. (1986). The generalized Mantel Haenszel statistic. *The American Statistician*, 40:106–108.

and English language status. If an item is not biased, then these comparisons should yield no difference in performance because the individuals being compared are already at the same level of overall achievement. On the other hand, if an item is biased against a particular sex or racial/ethnic group, there will be a difference in performance on that item, a difference that is inconsistent with overall test performance. The Mantel-Haenszel statistic indicates whether there are any statistically significant differences in performance; however, it does not show the magnitude of differences. Thus, following Dorans and Holland (1993)⁵, Mantel-Haenszel Delta DIF indices are calculated and items are classified into categories depending on DIF magnitudes.

- **IRT *a*-parameter**—Represents the degree to which the item differentiates between test takers with different abilities.
- **IRT *b*-parameter**—Represents the difficulty of an item. It sets the location of the inflection point of the item characteristic curve.
- **IRT *c*-parameter**—Represents the likelihood of correctly answering an item by guessing.
- **Item-Total Correlations**—Measures the correlation between the score on an item and IRT ability score derived from all operational items. Examples of item-total correlations are the point-biserial correlation, corrected-point biserial correlation, biserial correlation, poly-serial correlation, and the Pearson product moment correlation.
- ***p*-value**—A measure of student success on an item, equivalent to the mean score on the item divided by the total score points available for it. For multiple-choice and gridded-response items, this is the same as the percentage of students answering the item correctly. For multi-point items, it is the ratio of item's mean score to the number of points possible (analogous to the *p*-value).
- **Q1 Statistic**—Used as an index for how well an estimated item characteristic curve matches to observed item responses. Low values indicate good fit of an item to the observed responses. The ZQ1, an adjustment of the Q1 statistic, is used for analysis purposes.
- **Q3 Statistic**—Used as an index for how well the construct (measured by all items in a test) is unidimensional. IRT assumes that there is a strong, single construct that underlies the performance of all items. Low values of Q3 statistics across item pairs refer to the existence of a single construct.

Reliability Measures

Standard Error of Measurement (SEM), Marginal Reliability Index, Cronbach's Alpha—In statistical terms, reliability is a ratio of the variation in true achievement (that the test seeks to estimate) to variation in observed test scores, which are subject to error. If the error is minimal, the ratio will be close to one, and the test can be said to be reliable. The review of statistical characteristics is based on three indicators of reliability: conditional standard error of measurement, marginal reliability, and Cronbach's alpha. The Standard Error of Measurement

⁵ Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 35–66). Hillsdale, NJ: Lawrence Erlbaum Associates.

(SEM) describes the amount of error associated with the ability estimate. SEMs for the complete range of abilities are often represented graphically as conditional standard error curves to illustrate where the error is lowest. Typically, the error is lowest in the middle of the ability spectrum because there are more items associated with this level of abilities than at the extremes. Marginal reliability is a measure of the overall reliability of the test based on the average conditional SEM for all students. Cronbach’s alpha is a traditional measure of test reliability in which the degree of error is assumed to be the same at all levels of student achievement.

Achievement Level Classification Consistency and Accuracy—Consistency of classification is the agreement between classifications based on two equally difficult forms of the test. Accuracy of classification is the degree to which actual classifications agree with those that would be made on the basis of students’ true abilities, if they could be known. Three types of accuracy and consistency indices are estimated for the tests: overall, conditional-on-level, and by cut point. To describe consistency, these indices examine the agreement between actual performance and performance on a statistically modeled alternate and parallel test form. To describe accuracy, agreement between actual performance and a statistically constructed true score is examined.

Overall indices show the classification agreement grouped across all Achievement Levels, indices conditional-on-level outline the agreement at a selected Achievement Level, and indices by cut point score show the agreement around a single Achievement Level cut point.

Table 10: Statistical Analyses for Test Data and Indicators

Purpose	Indicator
Describe item difficulty	p -values, IRT b -parameters
Compare likelihood of success on item with likelihood of success on test	Item-total correlations, IRT a -parameters
Estimate of guessing	IRT c -parameters
Measure item fit to IRT model	Q1 (ZQ1) statistics
Measure test fit to IRT model (unidimensionality of achievement scale)	Q3 statistics
Identify bias	Differential Item Functioning (DIF) analysis (Mantel-Haenszel statistic, Generalized Mantel-Haenszel statistic)
Measure reliability	Conditional standard error of measurement, Marginal reliability index, Cronbach’s alpha
Verify Achievement Level classification accuracy and consistency	Indices of accuracy and consistency: overall, conditional-on-level, cut point

Appendix B: History, Requirements, and Uses

This appendix is organized as follows:

- History of Florida’s Statewide Assessments Program
- Overview of state statutes and the State Board of Education rules governing the statewide student assessment program
- State-required uses of statewide, standardized assessment results
- Primary statutory authority for the Statewide Student Assessment program

History of Florida’s Statewide Assessments Program

Florida’s focus on educational assessments and accountability began well before the first administration of the FCAT, which occurred in 1998. Key events in the state’s efforts to improve student achievement are described in the following condensed chronology. This summary outlines the origin of the student assessment and school accountability systems in Florida and how these efforts have changed over time.

1970s and 1980s

- Florida began administering statewide assessments in the 1970s.
- The nation’s first graduation test was authorized in 1976 and implemented first with the graduating class of 1983 (minimum competency test).

1990s

- The Florida Writing Assessment Program was administered for the first time to fourth-graders in 1992, expanding to eighth-graders in 1993 and tenth-graders in 1994.
- Identification of critically low-performing schools began in 1995, based on norm-referenced test scores in grades 4 and 8; writing scores in grades 4, 8, and 10; and results from the High School Competency Test in grade 11.
- FCAT was first administered in Reading (grades 4, 8, and 10) and Mathematics (grades 5, 8, and 10) in 1998.
- A–F School Grades were first issued in 1999, based on FCAT performance in the assessed grade levels and subjects, as well as additional indicators including dropout rates, attendance, and student discipline.
- Learning gains data were not yet available and were not part of School Grades from 1999 to 2001.

2000s

- FCAT Reading and Mathematics were expanded to grades 3–10 in 2001, allowing for the calculation of annual student learning gains.
- In 2002, the criteria for School Grades was expanded to include student learning gains and learning gains of the lowest-performing students (the Low 25%), with 50% of the grade based on student achievement and 50% based on learning gains.
- Passing the Grade 10 FCAT Reading and Mathematics exams (standards-based assessments) became a requirement for high school graduation beginning with the class of 2003.
- In 2003, FCAT Science was administered for the first time – once in elementary, once in middle, and once in high school.
- The 2006 Fall FCAT Retake was Florida’s first computer-based test administration and was administered to districts that volunteered to participate.
- In 2007, School Grades were expanded to include science performance and the learning gains of the Low 25% in mathematics.

2010s

- In 2010, School Grades for high schools were expanded to include acceleration, graduation rates, and college readiness.
- In 2011, Florida transitioned to FCAT 2.0, assessments developed to measure mastery of the Next Generation Sunshine State Standards.
- Florida EOC assessments began with Algebra 1 in 2011 and expanded to Geometry and Biology 1 in 2012, U.S. History in 2013, and Civics (middle school course) in 2014.
- In 2012, School Grades incorporated performance from FCAT 2.0 and EOC assessments for the first time.
- In 2014, School Grades included social studies achievement for the first time (U.S. History).
- In 2014–2015, Florida implemented the Florida Standards Assessments (FSA) in ELA and Mathematics, which measured mastery of the Florida Standards.
- In 2015, Informational Baseline School Grades were issued incorporating FSA performance for the first time, as well as a middle school acceleration component, under a new, simplified school grading model.
- In 2016, FSA Learning Gains components were included in School Grades since two years of FSA data were available.

2020s

- In 2020, the B.E.S.T. Standards for ELA and Mathematics were adopted by the State Board of Education.
- In the 2022–23 school year, assessments aligned to the B.E.S.T. Standards (FAST ELA Reading and Mathematics and B.E.S.T. Algebra 1 and Geometry EOC assessments) were administered for the first time.

Overview of State Statute and State Board of Education Rules Governing the Statewide Student Assessment Program

Table 11 provides an overview of the Florida Statutes and State Board of Education rules that govern Florida’s statewide assessments system, including their uses. Following the table, additional detail is provided regarding the required accountability uses for students, teacher, schools, and districts. Other assessment-related statutes and rules are listed in Table 12.

Table 11: Statewide, Standardized Assessments Statutes and Rules

Assessment	Assessment Citation	Required Use	Required Use Citation
Statewide Assessments Program (FAST, Science, EOCs, FAA)	s. 1008.22, F.S. Rule 6A-1.09422, F.A.C. Rule 6A-1.0943, F.A.C. Rule 6A-1.09432, F.A.C.	Measurement of student performance	s. 1012.34, F.S.
		Middle grades promotion	s. 1003.4156, F.S.
		Third grade retention; mid-year promotion; student progression; remedial instruction; reporting requirements	s. 1008.25, F.S. Rule 6A-1.094221, F.A.C. Rule 6A-1.094222, F.A.C.
		Differentiated accountability	s. 1008.33, F.S. Rule 6A-1.099811, F.A.C.
		School and district accountability	s. 1008.34, F.S. Rule 6A-1.09981, F.A.C.
		School improvement rating	s. 1008.341, F.S. Rule 6A-1.099822, F.A.C.
		Opportunity Scholarship	s. 1002.38, F.S.
		Standard high school diploma	s. 1003.4282, F.S.
		EOC assessments as 30% of course grade	s. 1003.4282, F.S.
		Concordant and comparative scores for graduation	Rule 6A-1.09422, F.A.C.

Assessment	Assessment Citation	Required Use	Required Use Citation
Statewide Assessments Program (FAST, Science, EOCs, FAA)	s. 1008.22, F.S. Rule 6A-1.09422, F.A.C. Rule 6A-1.0943, F.A.C. Rule 6A-1.09432, F.A.C.	Uniform assessment calendar	Rule 6A-1.094224, F.A.C.
		ESE graduation requirements and waiver of results	Rule 6A-1.09963, F.A.C.
Coordinated Screening and Progress Monitoring (FAST)	s. 1008.25, F.S. Rule 6A-1.09422, F.A.C.	VPK accountability	s. 1002.67, F.S. s. 1002.68, F.S. Rule 6M-8.620, F.A.C.
		Council for Early Grade Success	s. 1008.2125, F.S.
		Statewide, standardized assessments	s. 1008.22, F.S.
		Kindergarten screening	
FCLE	s. 1003.4282, F.S.	Postsecondary civic literacy assessment	s. 1007.25, F.S.
Department of Juvenile Justice (DJJ) Assessment – measures student learning gains and student progress while a student is in a juvenile justice education program	s. 1003.52, F.S. Rule 6A-6.05281, F.A.C.	DJJ accountability	
ACCESS for ELLs, Kindergarten ACCESS for ELLs, Alternate ACCESS – English language acquisition assessments to determine ESOL program placement or exit eligibility	s. 1003.56, F.S.	English for Speakers of Other Languages (ESOL) Exit	Rule 6A-6.0902, F.A.C. Rule 6A-6.0903, F.A.C. Rule 6A-6.09021, F.A.C. Rule 6A-6.0909, F.A.C.
Preliminary SAT (PSAT)/Preliminary ACT (PLAN) – administered by each public high school to all grade 10 students, though the parent has the opportunity to exempt his/her child from PSAT/PLAN	s. 1007.35, F.S.	Comparative score for graduation	Rule 6A-1.09422, F.A.C.
		Inform course placement	

Assessment	Assessment Citation	Required Use	Required Use Citation
Nationally recognized high school assessments (ACT, SAT, CLT) – administered by each public high school to grade 11 students; districts determine which of the three assessments to administer	s. 1008.22, F.S.	Concordant and comparative scores for graduation	Rule 6A-1.09422, F.A.C.
		Inform course placement; provide secondary opportunities	
National Assessment of Educational Progress (NAEP) – administered to a sample of students in selected grade levels every other year	s. 1008.22, F.S.	National and state comparisons	
Postsecondary Education Readiness Test (PERT) – a common placement test for determining readiness for college instruction in Florida	s. 1008.30, F.S.	Comparative score for graduation	Rule 6A-1.09422, F.A.C.
		Inform course placement	

Table 12: Other Related Statutes and Rules

Topic	Statute(s)	Rule(s)
Extraordinary Exemption/Medical Complexity Exemption for Students with Disabilities	s. 1008.212, F.S.	Rule 6A-1.0943, F.A.C.
Accommodations for Students with Disabilities	s. 1008.22, F.S.	Rule 6A-1.0943, F.A.C.
English Language Learners	s. 1008.22, F.S.	Rule 6A-1.09432, F.A.C. Rule 6A-6.0903, F.A.C.
Confidentiality of Assessment Instruments	s. 1008.23, F.S.	Rule 6A-10.042, F.A.C.
Test Security	s. 1008.24, F.S.	Rule 6A-10.042, F.A.C.
Family Empowerment Scholarship	s. 1002.394, F.S.	
Florida Tax Credit Scholarship Program	s. 1002.395, F.S.	
Home Education Programs	s. 1002.41, F.S.	

State-Required Uses of Statewide, Standardized Assessment Results

Accountability for Students

- A student must earn a Level 2 (out of 5) on the grade 3 ELA assessment to be promoted to grade 4.
 - Good cause exemptions are provided by law.
- A student must pass the grade 10 ELA Assessment to graduate from high school with a standard diploma.
 - Students may satisfy this requirement by earning a concordant score.
- A student must pass the Algebra 1 EOC Assessment to graduate from high school with a standard diploma.
 - Students may satisfy this requirement by earning a comparative score.
- A student enrolled in a course with a statewide, standardized EOC must take the assessment, and the results must count as 30% of the student’s course grade.
- A student who does not meet the required levels of performance on the assessment must be provided with additional diagnostic assessments and must participate in progress monitoring throughout the year.

Accountability for Schools and Districts

- The achievement and learning gains of students on the statewide, standardized assessments are used to determine school grades, district grades, and school improvement ratings for alternative and ESE center schools.
- Schools identified as schools in need of improvement based on student performance must provide progress monitoring.

Primary Statutory Authority for the Statewide Student Assessment Program

The primary statutory authority that addresses Florida’s statewide student assessment system is s. 1008.22, F.S. Per this statute, and as described in earlier sections, the primary purpose of the student assessment program is to provide student academic achievement and learning gains data to students, parents, teachers, school administrators, and district staff. Also per statute, assessment data are to be used by districts to improve instruction; by students, parents, and teachers to guide learning objectives; by education researchers to assess national and international education comparison data; and by the public to assess the cost benefit of the expenditure of taxpayer dollars. As with a number of other statutes, s. 1008.22, F.S., requires the State Board of Education to adopt rules to further define provisions of the statute.

Key components of s. 1008.22, F.S.:

- Participation in the statewide assessments program is **mandatory** for all students and for all districts, except in rare instances as noted below.

- The statewide, standardized assessment program must be aligned to the state content standards and must be administered in the following subjects and grade levels:
 - ELA grades 3–10,
 - Mathematics grades 3–8,
 - Science grades 5 and 8, and
 - EOC assessments in Algebra 1, Biology 1, Civics, Geometry, and U.S. History.
- Establishes the grades 3–10 ELA and grades 3–8 mathematics assessments as the coordinated screening and progress monitoring system described in section [1008.25](#), F.S.
- Requires five achievement levels for all statewide assessments, with level 1 being the lowest, level 5 being the highest, and level 3 indicating on-grade-level performance.
- Sets requirements for testing schedules, earliest test administration dates, and required reporting dates.
- Includes provision for the Florida Alternate Assessment (FAA) to be administered to certain students with disabilities, as determined by an Individual Education Plan (IEP) team.
- Includes provision for exemption from participation in statewide assessments for certain students with extraordinary circumstances or medical complexities.
- Provides the opportunity for students to meet graduation requirements through the use of concordant scores for grade 10 ELA and comparative scores Algebra 1.

Copyright Statement for This Office of Assessment Publication

Authorization for reproduction of this document is hereby granted to persons acting in an official capacity within the Uniform System of Public K–12 Schools as defined in Section 1000.01(4), Florida Statutes. The copyright notice at the bottom of this page must be included in all copies.

All trademarks and trade names found in this publication are the property of their respective owners and are not associated with the publishers of this publication.

This publication is provided by the Florida Department of Education to Florida public schools free of charge and is not intended for resale.

Permission is **not** granted for distribution or reproduction outside of the Uniform System of Public K–12 Schools or for commercial distribution of the copyrighted materials without written authorization from the Florida Department of Education. Questions regarding use of these copyrighted materials should be sent by email to **Assessment@fldoe.org** or by postal mail to the following address:

Bureau of K–12 Student Assessment
Florida Department of Education
325 West Gaines Street
Tallahassee, Florida 32399–0400

Copyright © 2024
State of Florida
Department of State