

STATEWIDE ASSESSMENT PROGRAM INFORMATION GUIDE

2019-2020



FLORIDA DEPARTMENT OF
EDUCATION
fldoe.org

Updated September 9, 2019

Table of Contents

Table of Tables	iv
1.0 Introduction.....	1
2.0 Test Content and Format	3
2.1. Operational, Field-Test, and Anchor Items.....	3
2.1.1. Test Design	3
2.1.2. Universal Design.....	3
2.1.3. Cognitive Complexity/Depth of Knowledge	4
3.0 Test Development and Construction.....	21
3.1. Developing Items	22
3.1.1. Item Writing.....	22
3.1.2. Committee Reviews	23
3.1.3. Field Testing	24
3.1.4. Statistical Review.....	24
3.1.5. Test Construction.....	25
3.1.6. Operational Testing.....	26
4.0 Test Administration.....	34
4.1. Schedules	34
4.2. Paper-Based Testing (PBT)	34
4.3. Computer-Based Testing (CBT).....	35
4.4. Timing.....	35
4.5. Personnel.....	37
4.6. Test Security	37
4.7. Administration Procedures.....	38
4.8. Accommodations	40
4.9. Test Invalidation	40
5.0 Scoring the Assessments	41
5.1. Standard Setting.....	41
5.2. Scoring Different Item Types	41
5.2.1. Machine/Electronic Scoring.....	41
5.2.2. Handscoring – FSA ELA Writing.....	42

5.2.3. Reported Scores	44
5.2.4. IRT Scoring.....	45
5.2.5. Process	46
6.0 Reporting Results.....	48
6.1. Reports	48
6.2. Florida’s PK–12 Education Information Portal	50
6.3. Florida School, District, and State Report Cards	50
7.0 Glossary	51
8.0 Guide to Useful Resources.....	57
8.1. ELA Writing Resources	57
8.2. Fact Sheets	57
8.3. Graduation Requirements	57
8.4. Practice Tests	57
8.5. Technical Reports	58
8.6. Test Administration Manuals.....	58
8.7. Test Design Summaries	58
8.8. Test Item Specifications.....	58
8.9. Schedules	59
8.10. Standard Setting.....	59
8.11. Aggregate Assessment Results	59
8.12. Understanding Reports.....	59
8.13. Useful Links.....	59
Appendix A: Statistical Indicators Used in Data Analysis.....	61
Indicator Definitions	61
Reliability Measures	62
Appendix B: History, Requirements, and Uses.....	64
History of Florida’s Statewide Assessment Program	64
1970s and 1980s.....	64
1990s	64
2000s	65
2010s	65
Overview of State Statute and State Board of Education Rules Governing the Statewide Student Assessment Program	66
State-Required Uses of Statewide, Standardized Assessment Results	68

Accountability for Students	68
Accountability for Schools and Districts	69
Primary Statutory Authority for the Statewide Student Assessment Program	69



Table of Tables

Table 1: FSA Percentage of Points by Depth of Knowledge Level	4
Table 2: NGSSS Percentage of Points by Cognitive Complexity Level	4
Table 3: Grade 4 Text-Based Writing.....	6
Table 4: Approximate Word Count Range for Text or Text Set	6
Table 5: Percentages of FSA Reading Text Types by Grade Band.....	7
Table 6: FSA ELA Reading Percentage of Points by Reporting Category	8
Table 7: Grade 3 FSA Mathematics Percentage of Points by Reporting Category	9
Table 8: Grade 4 FSA Mathematics Percentage of Points by Reporting Category	10
Table 9: Grade 5 FSA Mathematics Percentage of Points by Reporting Category	10
Table 10: Grade 6 FSA Mathematics Percentage of Points by Reporting Category	11
Table 11: Grade 7 FSA Mathematics Percentage of Points by Reporting Category	11
Table 12: Grade 8 FSA Mathematics Percentage of Points by Reporting Category	12
Table 13: FSA Algebra 1 Percentage of Points by Reporting Category.....	12
Table 14: FSA Geometry Percentage of Points by Reporting Category	13
Table 15: Statewide Science Percentage of Points by Reporting Category.....	17
Table 16: NGSSS Biology 1 Percentage of Points by Reporting Category	18
Table 17: NGSSS Civics Percentage of Points by Reporting Category	19
Table 18: NGSSS U.S. History Percentage of Points by Reporting Category	20
Table 19: FSA English Language Arts – Writing.....	35
Table 20: FSA English Language Arts – Reading.....	36
Table 21: FSA English Language Arts – Mathematics	36
Table 22: FSA End-of-Course Assessments.....	36
Table 23: NGSSS Assessments – Reading Retake and Science	36
Table 24: NGSSS Assessments – End-of-Course Assessments	36
Table 25: Grade 10 ELA Passing Score by Year.....	44
Table 26: Algebra 1 EOC Passing Score by Year	44
Table 27: Statistical Analyses for Test Data and Indicators	63
Table 28: Statewide, Standardized Assessments Statutes and Rules.....	66

1.0 Introduction

The primary purpose of Florida's K–12 statewide assessment program is to measure students' achievement of Florida's education 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 educational 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.

The *Statewide Assessment Program Information Guide* provides information about the Florida Standards Assessments (FSA) and the Next Generation Sunshine State Standards (NGSSS) Assessments, including 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 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 measurement practices. 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.

The FSA and NGSSS assessments measure student achievement of the standards contained in the Florida Standards and the Next Generation Sunshine State Standards, respectively. Florida's educational standards were developed with the goal of providing all students with an education based on high expectations. The statewide assessment program also provides feedback and accountability indicators to Florida educators, policy makers, students, and other citizens. This guide contains information about the FSA and NGSSS assessments only; the FDOE website contains information about the [Florida Standards Alternate Assessment \(FSAA\)](#), [ACCESS for ELLs](#), and the [National Assessment for Educational Progress \(NAEP\)](#).

It is important that 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 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 assessment 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 and how to assess them on the 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.

2.0 Test Content and Format

The FSA and NGSSS assessments are criterion-referenced tests that are intended to measure whether students have made progress on the English Language Arts Florida Standards, the Mathematics Florida Standards, the NGSSS Science Standards, and the NGSSS Social Studies Standards. Statewide assessments are constructed to meet rigorous 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 both computer-based test (CBT) format and paper-based test (PBT) format, and the item types listed in this section vary based on subject and delivery mode.

2.1. Operational, Field-Test, and Anchor Items

When taking statewide assessments, all students of the same grade level respond to a common set of items on each test. These common items are called operational items and count toward students' scores. 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.

Anchor items are those which have appeared on a given test in previous years and are used to ensure that the scores on that test can be equated or made comparable from year to year.

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 section lists the range of operational items from each reporting category that is required on each test form. This document guides item selection and test construction for FSA and NGSSS assessments to ensure that reporting categories and standards are correctly represented on each test form.

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

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

2.1.2. Universal Design

The application of universal design principles helps develop assessments that are usable to the greatest number of test takers, including students with disabilities and nonnative speakers of

English. To support the goal of providing access to all students, the test maximizes readability, legibility, and compatibility with accommodations, and test development 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, type size, line length, spacing, and graphics.

2.1.3. Cognitive Complexity/Depth of Knowledge

Statewide assessment items are classified using a model with origins in the works of Dr. Norman Webb¹ on depth of knowledge (DOK). With this system, items are classified on the cognitive demand inherent in the test item, not on assumptions about the student's approach to the item. The three categories—referred to as *DOK Level 1*, *DOK Level 2*, and *DOK Level 3* for FSA assessments, and low complexity, moderate complexity, and high complexity for NGSSS assessments—form an ordered description of the cognitive demands an item makes on a student. Items at the low level of complexity require a simple skill, such as locating details in a text or solving a one-step problem. At the moderate level, an item can ask the student to summarize a passage or retrieve information from a graph and use it to solve a problem. At the high level, an item may require a student to analyze cause-and-effect relationships or justify a solution to a problem. The distinctions made in item complexity are intended to provide a balance across the tasks administered at each grade level. The range of the percentage of points in each complexity level is listed for each assessment in the test blueprints and in the tables below.

Table 1: FSA Percentage of Points by Depth of Knowledge Level

Grade/Subject	DOK Level 1	DOK Level 2	DOK Level 3
Grades 3–10 ELA	10%–20%	60%–80%	10%–20%
Grades 3–8 Mathematics	10%–20%	60%–80%	10%–20%
Algebra 1 and Geometry	10%–20%	60%–80%	10%–20%

Table 2: NGSSS Percentage of Points by Cognitive Complexity Level

Grade/Subject	Low	Moderate	High
Grades 5 and 8 Science	10%–20%	60%–80%	10%–20%
Biology 1	10%–20%	60%–80%	10%–20%
Civics	15%–25%	45%–65%	15%–25%
U.S. History	20%–30%	45%–65%	15%–25%

¹ Webb, Norman L. and others. “Web Alignment Tool.” 24 July 2005. Wisconsin Center for Education Research. University of Wisconsin-Madison. 2 Feb 2006. <http://wat.wceruw.org/index.aspx>

English Language Arts (ELA)

In grade 3, students participate in FSA ELA Reading. In grades 4 through 10, the FSA ELA test includes two components, which are combined to provide a whole-test FSA ELA scale score: (1) A text-based Writing component in which students respond to one writing task by writing an essay, and (2) A reading and language component in which students respond to reading passages and other texts; in grades 7–10, students also respond to audio and multimedia content. Writing and Reading component item responses contribute to an overall ELA score. In this document, the term ELA is used when referring to the combined Reading and Writing assessments; ELA Reading is used when referring to only the Reading test form or items; and ELA Writing is used when referring only to the text-based Writing task.

ELA Writing

The Writing component of the grades 4–10 ELA assessments consists of one text-based constructed-response item (students read a variety of texts and respond to a prompt). There is no Writing component at grade 3. The rubrics used for the scoring of the FSA Writing component are based on the Writing and Language strands of the Language Arts Florida Standards.

To offer students a variety of texts on the FSA ELA Writing tests, authentic and copyrighted passages and articles appear as they were originally published, as requested by the publisher and/or author. While these real-world examples do not always adhere to strict style conventions and/or grammar rules, inconsistencies among passages should not detract from students' ability to understand and respond to the text-based writing task.

Once students read the provided texts, they respond to a prompt. There are two possible writing modes at each grade level. For grades 4 and 5, the prompt will ask for the writer's opinion or ask the writer to inform/explain. For grades 6–10, the prompt will ask the writer to provide an argument or to inform/explain. Students will draw upon the texts to provide evidence and information to support their claims or explanations.

Grades 4–6 students take a paper-based ELA Writing test. They read the texts and prompt in their test and answer books and provide a handwritten response on up to three lined pages.

Grades 7–10 students take a computer-based ELA Writing test. They read the texts and prompt in a secure, online environment. They then provide a typed response in a text box.

While the ELA Writing component is one category that contributes to the overall ELA score (Text-Based Writing), there are three domain scores awarded for each response. The sum of these subscores is the total score for the writing reporting category (10 raw score points).

The three domains for each grade level are:

- Purpose, Focus, and Organization (4 points)
- Evidence and Elaboration (4 points)
- Conventions of Standard English (2 points)

Table 3: Grade 4 Text-Based Writing Scoring Example shows a description of a Grade 4 student response that scored 8 points out of 10 possible points.

Table 3: Grade 4 Text-Based Writing Scoring Example

Purpose, Focus, and Organization (4 points possible)	Evidence and Elaboration (4 points possible)	Conventions of Standard English (2 points possible)
4 out of 4 possible points. The response is fully sustained and consistently focused within the purpose, audience, and task; it has a clearly stated controlling idea/opinion and effective organizational structure creating coherence and completeness.	3 out of 4 possible points. The response provides adequate support/evidence for the controlling idea/writer’s opinion that includes the use of sources, facts, and details.	1 out of 2 possible points. The response demonstrates a partial command of basic conventions. The response may include various errors in usage and inconsistent use of correct punctuation, capitalization, sentence formation, and spelling.

More detailed information may be found in the ELA Writing [scoring samplers and rubrics](#) on the FSA Portal.

ELA Reading

FSA ELA Reading is based on the standards found in the Reading, Literature, and Language strands of the English Language Arts Florida Standards. FSA ELA Reading employs a wide variety of written material to assess students’ reading comprehension as defined in the Florida Standards. Each FSA ELA Reading assessment is composed of approximately 6–8 reading passages with sets of 6–11 items based on each passage. There are two types of reading passages: informational and literary. In order to assess student mastery of grammar and standard English conventions, students are also asked to evaluate and correct grade-level errors in a “draft” essay.

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. Most passages are selected from published sources, although some may be written expressly for the FSA. As students progress beyond the early grades, they will read informational text with increasing frequency in and out of school. The percentage of informational text students will encounter on the FSA also increases as they progress through grades. Likewise, the range of words per passage increases across grade levels.

Table 4 suggests an approximate word count range for a text or text set.

Table 4: 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

Grade	Range of Number of Words
10	350–1350

Table 5 shows the percentages of FSA Reading text types by grade band.

Table 5: Percentages of FSA Reading Text Types by Grade Band

Grade Band	Literary	Informational
3–5	50%	50%
6–8	40%	60%
9–10	30%	70%

For all grade levels tested, FSA ELA assesses what students know and are able to do in the broad reporting categories listed below. More detailed explanations can be found on the [FDOE Standard Setting page](#). The difficulty of the concepts assessed on FSA ELA progresses systematically from grade to grade, as does the complexity of the text presented to the student at each grade level.

Grade 3

- **Key Ideas and Details**
Students read closely to comprehend, analyze, and summarize essential information and concepts, referencing evidence from the text to support inferences and conclusions.
- **Craft Structure**
Students interpret literal and nonliteral meanings of words/phrases. They determine how text structures and text features impact meaning. They distinguish personal point of view from that of the narrator or author.
- **Integration of Knowledge and Ideas**
Students integrate and analyze content presented in diverse media formats. They analyze treatment of similar themes or topics.
- **Language and Editing**
Students demonstrate command of the conventions of Standard English grammar, usage, capitalization, punctuation, and spelling. Items may ask the student to evaluate and correct errors which focus on grammar and usage or capitalization, punctuation, and spelling. Items should assess on-grade-level errors; however, once a language standard is introduced, grade-appropriate items may be written to assess continued mastery of standard conventions of English.

Grades 4–5

- **Key Ideas and Details**
Students read closely to comprehend, analyze, and summarize essential information and concepts, citing textual evidence to support inferences and conclusions.
- **Craft and Structure**
Students interpret connotative and figurative meanings of words/phrases. They analyze how text structures and text features impact the text. They determine the effects of point of view or purpose.
- **Integration of Knowledge and Ideas**

Students integrate and evaluate content presented in diverse media formats. They analyze the treatment of similar themes or topics and how the author uses reasons and evidence to support points.

- **Language and Editing**

Students demonstrate command of the conventions of Standard English grammar, usage, capitalization, punctuation, and spelling. Items may ask the student to evaluate and correct errors which focus on grammar and usage or capitalization, punctuation, and spelling. Items should assess on-grade-level errors; however, once a Language Standard is introduced, grade-appropriate items may be written to assess continued mastery of standard conventions of English.

- **Text-Based Writing**

Students draw relevant evidence from various texts to support a claim or controlling idea. They produce clear and coherent writing with development, organization, and style appropriate to task, purpose, and audience.

Grades 6–10

- **Key Ideas and Details**

Students read closely to understand information. They cite textual evidence to support inferences/conclusions. They analyze development and interaction of central ideas, themes, individuals, events, or supporting ideas. They summarize key concepts.

- **Craft and Structure**

Students interpret connotative and figurative meanings of words/phrases. They analyze how word choice affects meaning/ tone and how text structures impact the text. They determine the effects of point of view or purpose.

- **Integration of Knowledge and Ideas**

Students integrate and evaluate content presented in diverse media formats. They evaluate arguments for claims, validity, relevance, and sufficient evidence. They analyze treatment of similar themes or topics.

- **Language and Editing**

Students demonstrate command of the conventions of Standard English grammar, usage, capitalization, punctuation, and spelling. Items may ask the student to evaluate and correct errors which focus on grammar and usage or capitalization, punctuation, and spelling. Items should assess on-grade-level errors; however, once a Language Standard is introduced, grade-appropriate items may be written to assess continued mastery of standard conventions of English.

- **Text-Based Writing**

Students draw relevant evidence from various texts to support a claim or controlling idea. They produce clear and coherent writing with development, organization, and style appropriate to task, purpose, and audience.

Table 6: FSA ELA Reading Percentage of Points by Reporting Category

Grades	Key Ideas and Details	Craft and Structure	Integration of Knowledge and Ideas	Language and Editing
3–10	15%–25%	25%–35%	20%–30%	15%–25%

Mathematics

FSA Mathematics is based on the Mathematics Florida Standards and the Florida Course Descriptions. FSA Mathematics is administered to students in grades 3–8, and two mathematics FSA EOCs, Algebra 1 and Geometry, are administered to students enrolled in and completing the respective course (or an equivalent course). See the [FSA EOC 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 Standards Assessments \(FSA\) Mathematics Assessments](#) document.

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

Grade 3

- Operations, Algebraic Thinking, and Numbers in Base Ten**
 Students represent and solve problems involving multiplication and division. They understand properties of multiplication and the relationship between multiplication and division. They multiply and divide within 100. They solve problems involving the four operations, and identify and explain patterns in arithmetic. They use place value understanding and properties of operations to perform multi-digit arithmetic.
- Numbers and Operations—Fractions**
 Students develop understanding of fractions as numbers.
- Measurement, Data, and Geometry**
 Students solve problems involving measurement and estimation of intervals of time, liquid volumes, and masses of objects. They represent and interpret data. They understand concepts of area and relate area to multiplication and addition. They recognize perimeter as an attribute of plane figures and distinguish between linear and area measures. They reason with shapes and their attributes.

Table 7: Grade 3 FSA Mathematics Percentage of Points by Reporting Category

Operations, Algebraic Thinking, and Numbers in Base Ten	Numbers and Operations—Fractions	Measurement, Data, and Geometry
48%	17%	35%

Grade 4

- Operations and Algebraic Thinking**
 Students use the four operations with whole numbers to solve problems. They gain familiarity with factors and multiples. They generate and analyze patterns.
- Numbers and Operations in Base Ten**
 Students generalize place value understanding for multi-digit whole numbers. They use place value understanding and properties of operations to perform multi-digit arithmetic.

- **Numbers and Operations—Fractions**

Students extend understanding of fraction equivalence and ordering. They build fractions from unit fractions by applying and extending previous understandings of operations on whole numbers. They understand decimal notation for fractions and compare decimal fractions.

- **Measurement, Data, and Geometry**

Students solve problems involving measurement and conversion of measurements from a larger unit to a smaller unit. They represent and interpret data. They understand concepts of angle and measure angles. They draw and identify lines and angles and classify shapes by properties of their lines and angles.

Table 8: Grade 4 FSA Mathematics Percentage of Points by Reporting Category

Operations and Algebraic Thinking	Numbers and Operations in Base Ten	Numbers and Operations—Fractions	Measurement, Data, and Geometry
21%	21%	25%	33%

Grade 5

- **Operations, Algebraic Thinking, and Fractions**

Students write and interpret numerical expressions. They analyze patterns and relationships. They use equivalent fractions as a strategy to add and subtract fractions. They apply and extend previous understandings of multiplication and division to multiply and divide fractions.

- **Numbers and Operations in Base Ten**

Students understand the place value system. They perform operations with multi-digit whole numbers and decimals to hundredths.

- **Measurement, Data, and Geometry**

Students convert like measurement units within a given measurement system. They represent and interpret data. They understand concepts of volume and relate volume to multiplication and addition. They graph points on the coordinate plane to solve real-world and mathematical problems. They classify two-dimensional figures into categories based on their properties.

Table 9: Grade 5 FSA Mathematics Percentage of Points by Reporting Category

Operations, Algebraic Thinking, and Fractions	Numbers and Operations in Base Ten	Measurement, Data, and Geometry
39%	28%	33%

Grade 6

- **Ratio and Proportional Relationships**

Students understand ratio concepts and use ratio reasoning to solve problems.

- **Expressions and Equations**

Students apply and extend previous understandings of arithmetic to algebraic expressions. They reason about and solve one-variable equations and inequalities. They represent and analyze quantitative relationships between dependent and independent variables.

- **Geometry**
Students solve real-world and mathematical problems involving area, surface area, and volume.
- **Statistics and Probability**
Students develop understanding of statistical variability. They summarize and describe distributions.
- **The Number System**
Students apply and extend previous understandings of multiplication and division to divide fractions by fractions. They compute fluently with multi-digit numbers and find common factors and multiples. They apply and extend previous understandings of numbers to the system of rational numbers.

Table 10: Grade 6 FSA Mathematics Percentage of Points by Reporting Category

Ratio and Proportional Relationships	Expressions and Equations	Geometry	Statistics & Probability	The Number System
15%	30%	15%	19%	21%

Grade 7

- **Ratio and Proportional Relationships**
Students analyze proportional relationships and use them to solve real-world and mathematical problems.
- **Expressions and Equations**
Students use properties of operations to generate equivalent expressions. They solve real-life and mathematical problems using numerical and algebraic expressions and equations.
- **Geometry**
Students draw, construct, and describe geometrical figures and describe the relationships between them. They solve real-life and mathematical problems involving angle measure, area, surface area, and volume.
- **Statistics and Probability**
Students use random sampling to draw inferences about a population. They draw informal comparative inferences about two populations. They investigate chance processes and develop, use, and evaluate probability models.
- **The Number System**
Students apply and extend previous understandings of operations with fractions to add, subtract, multiply, and divide rational numbers.

Table 11: Grade 7 FSA Mathematics Percentage of Points by Reporting Category

Ratio and Proportional Relationships	Expressions and Equations	Geometry	Statistics & Probability	The Number System
25%	21%	23%	16%	15%

Grade 8

- **Expressions and Equations**
Students work with radicals and integer exponents. They understand the connections between proportional relationships, lines, and linear equations.
- **Functions**
Students define, evaluate, and compare functions. They use functions to model relationships between quantities.
- **Geometry**
Students understand congruence and similarity using physical models, transparencies, or geometry software. They understand and apply the Pythagorean Theorem. They solve real-world and mathematical problems involving volume of cylinders, cones, and spheres.
- **Statistics and Probability and the Number System**
Students investigate patterns of association in bivariate data. They know that there are numbers that are not rational and approximate them by rational numbers.

Table 12: Grade 8 FSA Mathematics Percentage of Points by Reporting Category

Expressions and Equations	Functions	Geometry	Statistics & Probability and The Number System
30%	25%	27%	18%

Algebra 1

- **Algebra and Modeling**
Students perform operations on polynomials. They understand the relationship between zeros and factors of polynomials. They use mathematical structure of expressions. They create, solve, and reason with equations and inequalities. They choose and use appropriate mathematics to model situations.
- **Functions and Modeling**
Students understand the concept of a function. They interpret functions and key features in a context. They analyze and graph functions. They build a function that models a relationship. They construct linear, quadratic, and exponential functions. They solve problems using functions.
- **Statistics and the Number System**
Students extend the properties of exponents to rational exponents. They use properties of rational and irrational numbers. They summarize, represent, and interpret data for one- and two-variable data. They interpret linear models.

Table 13: FSA Algebra 1 Percentage of Points by Reporting Category

Algebra and Modeling	Functions and Modeling	Statistics and The Number System
41%	40%	19%

Geometry

- Congruence, Similarity, Right Triangles, and Trigonometry**
 Students understand congruence and similarity in terms of transformations. They prove and use geometric theorems. They demonstrate geometric constructions. They define trigonometric ratios. They solve problems involving right triangles. They use congruence and similarity criteria for triangles.
- Circles, Geometric Measurement, and Geometric Properties with Equations**
 Students prove and apply theorems about circles. They find arc lengths and areas of sectors. They derive the equation of a circle. They use coordinates to prove theorems and to solve problems algebraically. They explain and use volume formulas.
- Modeling with Geometry**
 Students apply geometric concepts in modeling situations.

Table 14: FSA Geometry Percentage of Points by Reporting Category

Congruence, Similarity, Right Triangles, and Trigonometry	Circles, Geometric Measurement, and Geometric Properties with Equations	Modeling with Geometry
46%	38%	16%

FSA ELA 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 FSA ELA Reading and Mathematics tests. If a CBT item type is modified in any way so that it may be presented on a paper-based form, that information is included in the descriptions below.

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. For paper-based assessments, this item type may be replaced with a modified version of the item that can be scanned and scored electronically, or replaced with another item type that assesses the same standard and can be scanned and scored electronically.

Grades 3–6 Paper-Based ELA Item Types

- Multiple-Choice:** Students select one correct answer from four answer choices.
- Editing Task Choice:** The student reads a short passage with underlined words and phrases. The student selects the correct word or phrase that should replace the underlined word or phrase from a set of options. One option will always be correct “as is.”
- 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.
- 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.

5. **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.
6. **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.
7. **Multimedia:** Enhanced content may include multimedia elements such as images, charts, graphics timelines, etc. 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.

Grades 3–6 Paper-Based Mathematics Item Types

1. **Multiple-Choice:** Students select one correct answer from four answer choices.
2. **Equation Editor/Gridded-Response:** Students enter a value into a grid. Grade 3 grids have number and fraction bar bubbles, grades 4 and 5 grids have number, fraction bar, and decimal point bubbles, and grade 6 grids have number, fraction bar, decimal point, and negative sign bubbles. For examples of grids on grades 7 and 8 and EOC Mathematics paper-based accommodations, see the paper-based practice tests on the FSA Portal.

	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$
0	0	0	0	0	0	0
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9

	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$
.
0	0	0	0	0	0	0
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9

-	-	-	-	-	-	-
	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$	$\frac{\Box}{\Box}$
.
0	0	0	0	0	0	0
1	1	1	1	1	1	1
2	2	2	2	2	2	2
3	3	3	3	3	3	3
4	4	4	4	4	4	4
5	5	5	5	5	5	5
6	6	6	6	6	6	6
7	7	7	7	7	7	7
8	8	8	8	8	8	8
9	9	9	9	9	9	9

3. **Editing Task Choice:** The student fills in a bubble to indicate the correct number, word, or phrase that should complete a phrase or sentence.
4. **Selectable Hot Text:** Excerpted sentences from the text are presented in this item type. The student fills in bubbles to indicate which sentences are correct.
5. **Multiselect:** The student is directed to select all of the 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.

6. **Matching:** This item type presents options in columns and rows. 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.

Grades 7–10 Computer-Based ELA and Mathematics Item Types

1. **Multiple-Choice:** Students select one correct answer from four answer choices.
2. **Editing Task Choice:** The student clicks a highlighted word or phrase, which reveals a drop-down menu containing options for correcting an error as well as the highlighted word or phrase as it is shown in the sentence to indicate that no correction is needed. The student then selects the correct word or phrase from the drop-down menu. For paper-based assessments, the item is modified so that it can be scanned and scored electronically. The student fills in a bubble to indicate the correct word or phrase.
3. **Hot Text**
 - a. **Selectable Hot Text:** Excerpted sentences from the text are presented in this item type. When the student hovers over certain words, phrases, or sentences, the options are highlighted. This indicates that the text is selectable (“hot”). The student can then click an option to select it. These items may have one or two parts. In a two-part hot text item, Part A may ask the student to make an analysis or an inference, and Part B may require the student to use the text to support the answer in Part A. In other cases, the two parts may function independently. For paper-based assessments, a “selectable” hot text item is modified so that it can be scanned and scored electronically. In this version, the student fills in a bubble to indicate a selection.
 - b. **Drag-and-Drop Hot Text:** Certain words, phrases, or sentences may be designated “draggable” in this item type. When the student hovers over these areas, the text is highlighted. The student can then click on the option, hold down the mouse button, and drag the option to a graphic organizer or other format. For paper-based assessments, drag-and-drop hot text items will be modified or replaced with another item type that assesses the same standard/reporting category and can be scanned and scored electronically.
4. **Open Response:** The student uses the keyboard to enter a response into a text field. These items may require entering a numerical value or can usually be answered in a sentence or two. For accommodated paper-based assessments, this item type may be replaced with another item type that assesses the same standard/reporting category and can be scanned and scored electronically.
5. **Multiselect:** The student is directed to select a specific number or all of the correct answers from among the options provided. These items are different from multiple-choice items, which allow the student to select only one correct answer. These items appear in both the online tests and paper-based accommodations.
6. **Evidence-Based Selected Response (ELA only):** 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. These items appear in the online and paper-based assessments.
7. **Graphic Response Item Display (GRID):** The student may select numbers, words, phrases, or images and use the drag-and-drop feature to place them into a graphic organizer or other

format. This item type may also require the student to use the point, line, or arrow tools to create a response on a graph. For paper-based assessments, this item type may be replaced with another item type that assesses the same standard/reporting category and can be scanned and scored electronically.

- 8. 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. The student is directed to click 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. These items appear in the online and paper-based assessments. For paper-based assessments, the item is modified so that it can be scanned and scored electronically. The student fills in a bubble to indicate the correct selection.
- 9. Multimedia (ELA only):** Technology-enhanced content may include multimedia elements such as audio clips, slideshows, or animations. 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 paper-based assessments, multimedia content may be modified or replaced by paper-based items or stimuli that assess the same reporting category.
- 10. Equation Editor (Mathematics only):** The student is presented with a toolbar that includes a variety of mathematical symbols that can be used to create a response. Responses may be in the form of a number, variable, expression, or equation, as appropriate to the test item. For paper-based assessments, this item type may be replaced with a modified version of the item that can be scanned and scored electronically or replaced with another item type that assesses the same standard and can be scanned and scored electronically.
- 11. Table Item (Mathematics only):** The student types numeric values into a given table. The student may complete the entire table or portions of the table depending on what is being asked. For paper-based assessments, this item type may be replaced with another item type that assesses the same standard and can be scanned and scored electronically.

[CBT and PBT ELA and Mathematics Practice Tests](#) are on the FSA Portal and include examples of all item types listed above. The [Practice Tests Guide](#) provides an overview of the item types that students may see on the practice tests and general guidelines on how to respond to items.

Science

Florida's Science assessments are based on the benchmarks found in the Science NGSSS and the Florida Course Descriptions. The Statewide Science Assessment is administered to students in grades 5 and 8, and the NGSSS Biology 1 EOC Assessment is administered to students enrolled in and completing the course (or an equivalent course). See [the NGSSS EOC fact sheet](#) for a list of courses.

The Grades 5 and 8 Statewide Science Assessments include multiple-choice items only. 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 15: Statewide Science Percentage of Points 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 NGSSS 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 [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 16: NGSSS Biology 1 Percentage of Points 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 NGSSS and the Florida Course Descriptions. The NGSSS Civics and U.S. History EOC assessments are administered to students enrolled in and completing the respective course (or an equivalent course). See the [NGSSS EOC fact sheet](#) for a list of courses.

Civics

The Civics EOC Assessment measures student achievement of the NGSSS in social studies, as outlined in the M/J Civics 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 [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 17: NGSSS Civics Percentage of Points by Reporting Category

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

U.S. History

The U.S. History EOC Assessment measures student achievement of the NGSSS in social studies, as outlined in the United States History 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 [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.

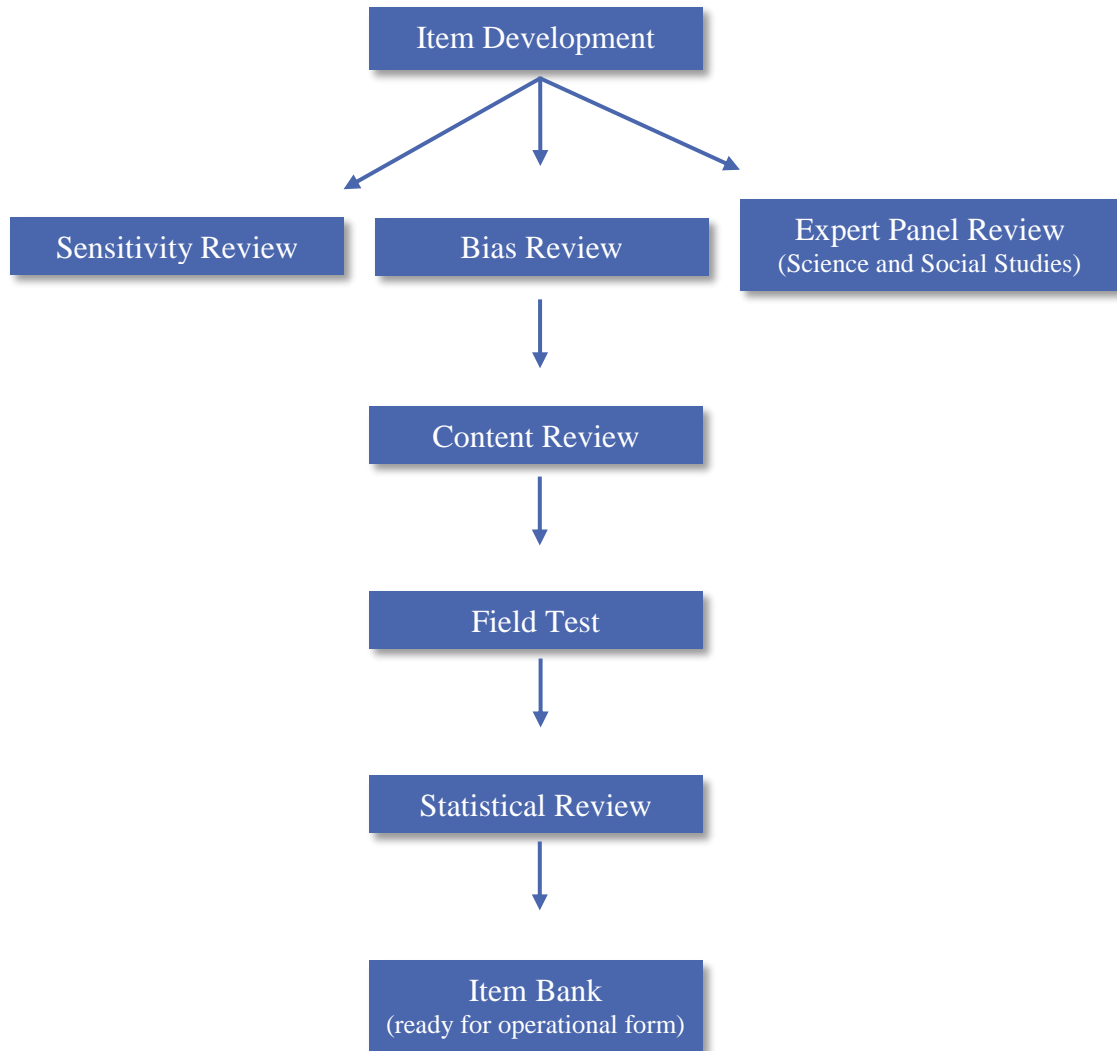
- **The United States and the Defense of the International Peace (1940–2010)**
Students understand and articulate the impact of issues related to World War II, the Cold War, the social revolutions of the late twentieth century, and the challenges of the early twenty-first century.

Table 18: NGSSS U.S. History Percentage of Points 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 and Construction

Developing statewide assessments to accurately measure achievement and accurately compare 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 and construction processes, 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 the test as an operational item, which counts toward a student's score.

The second key concept relates to the nature of the item writing and test construction processes. 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, they become part of the item bank. The item bank is a database of items serving as the source for constructing a test each year. The process of test construction involves selecting a set of items from the item bank that meets the established content and statistical guidelines of the test. The operational items on a test in any given year will likely have been written in another year and may appear on a test form several times before being retired.

3.1. Developing Items

There are six 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. Test Construction
6. Operational Testing

3.1.1. Item Writing

For each subject and grade level, criteria for item development are specified in the FSA and NGSSS Test Item Specifications. The specifications include the specific benchmarks, the types of items used, guidelines for the relative balance of topics, item formats and complexity levels, plus 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, writing samples, and reading passages.

Each year, for all grades and subjects, FDOE, Florida educators, and the contractor(s) 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 varied and often specialized backgrounds and abilities, and have teaching experience. All item writers are required to attend a training session that includes a review of item specifications, cognitive complexity levels, good item 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. Each item writer is given multiple opportunities to draft and evaluate items during training. After training, item writers are assigned to write and submit items for review. Items are reviewed and edited several times by FDOE 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.

Items for all four subject areas 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. They look for any items, prompts, graphics, or passages that might provide an advantage or disadvantage (unrelated to an understanding of the content) to a student with certain personal characteristics, such as those related to gender, race, ethnicity, religion, socioeconomic status, disability, or geographic region.

Similar to the Bias Review Committees, the Community Sensitivity Committees are made up of Florida citizens associated with a variety of organizations and institutions. Membership is drawn from statewide religious organizations, parent organizations, community-based 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 sensitivity are distinct from bias because 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. The Community Sensitivity Committees meet once or twice a year.

After each 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 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, evaluate the specified levels of cognitive complexity, are clearly worded, have only one correct answer (for multiple-choice items), and are grade-level or course appropriate. Committee members also recommend approval, modification, or rejection of the passages or items presented by FDOE. There are four Item Content Review Committees, one for each subject with grade-level subcommittees, which usually meet in the fall. Committee members for all four content areas serve on an ad hoc basis. Another reading committee meets only to review potential reading passages. 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 can 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. On a test with 45–65 items, most test forms will contain six to ten field-test items. In November/December of 2019, students in selected schools in the state will participate in a standalone FSA ELA Writing Field Test to replenish the text-based writing item bank. 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.

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 and gridded-response items must undergo a further review.

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 papers 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 will be reviewed by a Rangefinder Review Committee.

Because Florida will conduct an FSA ELA Writing Field Test in late 2019, a Prompt Review Committee met in Spring 2019 to select prompts for the field test. This review ensures that prompts selected for the field test employ clear wording, are of appropriate difficulty and interest level, and are unbiased. Participants included language arts teachers from the targeted grade levels, and school and district curriculum specialists.

Rubric Validation and Gridded-Response Adjudication Committees review all responses to field-tested gridded-response items to determine whether all possible correct answers have been included in the scoring key. Based on their input, FDOE establishes rules for how each constructed-response 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 constructed-response 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 an image of the item, the item statistics, and details about the item's location on the test form.

The statistical review of these items is conducted as an initial step of test construction. 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. Test Construction

Test construction is guided by a set of Test Construction Specifications, which are based on the Test Item Specifications, and other considerations such as statistical criteria. Because the Test Construction Specifications are used to develop a complete test for a single year, they include more detail about how standards are addressed and about statistical characteristics of items and the final test. The Test Construction Specifications are revised annually to guide the construction of each assessment. Because they contain very detailed information about the content and psychometric characteristics of the tests, the Test Construction Specifications are protected by test security statutes and are not available to the public.

During the summer months, prior to each test administration, FDOE uses the Test Construction Specifications to carefully select items for use on the statewide assessments in the upcoming school year. For ELA and Mathematics assessments, a single set of operational items is selected to which either field-test or anchor items are added to create the test forms for each subject and grade. For EOC assessments, several sets of operational items are selected to which either field-test or anchor items are added, and alternative test forms are created for each subject. Next, FDOE approves the basic components of the test through a series of reviews resulting in a final version of each assessment.

Test Construction Process

Each test form is carefully constructed using statistical programs and teams of content and psychometric experts from both FDOE and the contractor. These teams meet in person for three weeks each summer to select test forms for the following spring. The face-to-face interaction allows for staff to collaborate in a meaningful, effective way. The following steps describe how a form is constructed from its first version to the approved final product.

1. Contractor content staff select operational items that follow the test blueprint. The anchor items are selected first, and these items link student performance from one year to the next. Then the “core” items, which are the items that contribute to student scores, are selected next, and like the set of anchor items, follow the blueprint. The statistical qualities and content coverage of anchor sets and core item sets within a test are designed to be very similar to each other.
2. Initial versions of the form are provided to FDOE content experts for review. These content specialists collaborate with contractor content specialists to revise the forms and select replacement items, if needed. Both parties select items with respect to the statistical guidelines and the content and blueprint guidelines. If a content specialist has any concerns about blueprints or statistical summaries, it will be discussed with FDOE psychometricians during this stage. Once a form is approved by department content leads, it is sent for review by FDOE psychometricians.

3. The FDOE psychometric team evaluates the statistical properties of the constructed forms against the statistical targets outlined in the Test Construction Specifications. The information analyzed for each form includes desired average test difficulty, average item difficulty, target test information, standard error of measurement, and test characteristic curves. At the item level, each test item is reviewed for classical item difficulty, classical item discrimination, differential item function, item response theory, and fit statistics. More information on these topics can be found in [Appendix A](#). The proposed forms are then either returned to the content teams for suggested improvements, or are approved and forwarded to FDOE leadership for final review.
4. All proposed forms are reviewed by FDOE's leadership team to determine the suitability of the selected items and test forms as a whole. When evaluating any given form, leadership considers the diversity of topics, the projected level of difficulty, statistical summaries, adherence to the blueprint, and overall presentation to students. The leadership team is given the opportunity to approve proposed forms or return them with comments to the content team for further revision.
5. Once a form receives final sign-off, field-test items are selected for placement within certain test forms. Field-test items do not count toward student scores.

3.1.6. Operational Testing

Operational testing occurs when an assessment is administered in Florida public schools in the spring of each year. Because of the multi-step item development process and the use of the item bank, many operational items will have been written and reviewed at least two school years prior to appearing on the test.

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 both the items and the test as a whole meet established design and psychometric criteria, as the field-test results indicated they would.

Developing Tests

After committee reviews and field testing are completed, the process of selecting items to construct a test begins. The process of design and construction of each test form targets important goals but is also constrained by the realities of cost and time. Since the purpose of the assessments is to measure student achievement of Florida standards, items must have clear connections to those standards. To be of value, test scores must accurately represent students' abilities, requiring not only a large enough sample of student work—in this case, a sufficient number of items—but also items providing specific types of information about student achievement. Constructing a statewide assessment requires using the science of psychometrics. For example, statistical analyses are used to verify the quality of the individual items and the validity of the test as a whole. In addition, the need for comparable results from year to year requires that the test design maintains consistent content and statistical characteristics. The test

should be appropriate for Florida’s diverse student population and acceptable to all communities in Florida, while still providing an accurate assessment of the standards.

In order for the assessments to serve their various functions within the limitations placed upon them, very clear criteria and quality control measures are established for designing both test items and the test itself. The criteria and the quality control measures are partially based on the recommendations of the Technical Advisory Committee.

The next sections present descriptions of the desired characteristics of test items and the entire test, as well as the measures taken to ensure them. 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 the field test and again after each operational test to verify that the items performed as expected. Definitions for the terms throughout this section can be found at the end of the document 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. 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 (analogous 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 item characteristics to 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, it may be rejected from 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.

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 and corrected-point biserial correlations 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 percent. 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.

Instead of a four-choice item, the item essentially becomes a three- or two-choice item. To minimize guessing on a multiple-choice item, item writers and reviewers are instructed to design items with plausible distractors, but 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 guessing. Items with unusually high 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, that are unrelated to academic achievement.

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 who note any potential bias and give their comments to item reviewers. In some cases, items are eliminated from further consideration at this point.

In addition to the thorough reviews by the Bias and Community Sensitivity Review Committees, bias is also investigated via statistical analyses, named as 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 gender (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 contractors 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.

Item Fit to the IRT Model

Data analyses conducted after field testing and after operational testing generate IRT statistics for each item. These statistics are for 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 guessing (the c -parameter). They are used to establish a mathematical model that specifies the relation of student ability levels and item characteristics to a student's item response. This mathematical model, also named Item Response Function, is expected to fit the observed pattern of student responses.

Quality Assurance Measures—For each item, a statistic describing the quality of fit to the model is generated. This statistic is derived by estimating 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 fit values that indicate a good fit of the model. These standards are established in the Test Construction Specifications.

Characteristics of the Test

This section describes the desired characteristics of the assessment forms prepared annually. Each characteristic is followed by an explanation of the related quality assurance method.

Content Coverage (Content Validity)

The statewide assessments measure student success on the Florida Standards (ELA and Mathematics) and Next Generation Sunshine State Standards (Science and Social Studies) with a balance of emphasis among them. It is important that the assessments include items that collectively reflect the desired range of those standards. Results from a test that does not sufficiently sample the set of standards or the content domain will not provide an accurate measure of achievement in that subject area.

Quality Assurance Measures—Each year, test developers use the guidelines in the Test Construction Specifications to develop the assessments. This document specifies the number of items to address each standard and the percentage distribution of items across content strands or clusters. The Test Construction Specifications help FDOE's test developers ensure that the test forms reflect the range and balance of content specified in the set of standards used to define the subject area.

Test Difficulty

When all the items on a test are of the same level of difficulty, results tend to identify two groups of students: those who can correctly answer questions at the given difficulty level and those who cannot. It is more desirable that the items on a test address a range of knowledge of the content being assessed. When items represent a range of difficulty levels, it is much easier to identify students achieving at relatively higher levels (those who are able to correctly answer the most difficult items) and at relatively lower levels (those who are unable to correctly answer the easiest items). Generally speaking, a range of item difficulties allows creation of a scale of student achievement with useful information on students at all levels of achievement.

Quality Assurance Measures—Assuring the necessary range of item difficulties occurs mainly during test construction. In addition to selecting items for content coverage, test developers select items based on difficulty-related data gathered either from field tests or from operational use in previous years. The two indicators of item difficulty used in test construction (the items' p -values and IRT b -parameters) are the same as those used in item-level analysis. During test construction, test developers review both the p -values and b -parameters for all items to ensure distribution of item difficulties across all levels of achievement.

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 error. On a reliable test, the amount of error will be small. When there are sufficient numbers of test items 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.

Quality Assurance Measures—Virtually all of the steps in the test development process contribute in some way or another to minimize error and maximize the reliability of the assessments. In the process of test construction, test developers review the statistical data for items and generate three indicators of overall test reliability: Conditional Standard Error of Measurement (CSEM), IRT marginal reliability, and Cronbach's alpha reliability coefficient. These statistics and measures are reviewed in light of established guidelines before final approval. CSEM, test information curves, IRT marginal reliability, and Cronbach's alpha, reliability coefficients are all reviewed at test construction and after test administration.

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. 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, have confirmed that each test, as developed, fits the IRT model.

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 Item Response Theory (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. During test construction, item parameters are carefully reviewed to determine if an item is suitable to become an operational item. The parameters are recalculated after operational use and then used for live scoring.

Items differ in their difficulty such that the position of the point of inflection of this curve (the vertical line on the example on the previous page) 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)² is used to analyze multiple-choice items. The two-parameter logistic (2PL) model is a limited version of the 3PL model where the *c*-parameter (guessing) is set to zero. This model is used to analyze some single-point (correct or incorrect) technology enhanced items (e.g., open response). The Generalized Partial Credit Model (Muraki, 1997)³ is used to analyze multi-point items, such as writing prompts.

IRT item parameter estimates for all items on a test provide the means for determining scores of 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 likelihood score estimate for each student, based on their responses and item characteristics. Using the sophisticated IRT model is advantageous for large-scale testing programs, such as Florida's, because it helps create a stable scoring system when items included on the tests change from one year to the next.

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 representatives of universities and state agencies outside Florida. In addition, the psychometric advisors of FDOE's contractors participate in the committee meetings. Committee members assist the

² Lord, F.M. & Novick, M.R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.

³ Muraki, E. (1997). A generalized partial credit model: application of an EM algorithm. *Applied Measurement*, 7, 159–176.

DOE 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 once or twice a year.

4.0 Test Administration

After the test has been designed, items have been field-tested and approved, and test forms have been finalized, the next step is to administer the test to students. 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 for schools chosen for calibration subjects. The state-established testing windows for the Statewide Assessment Program are posted on the [FDOE website](#).

4.2. Paper-Based Testing (PBT)

In the 2019–2020 school year, the following assessments are PBT administrations:

- Grades 4–6 FSA ELA Writing
- Grades 3–6 FSA ELA Reading
- Grades 3–6 FSA Mathematics
- Grades 5 and 8 Statewide Science

The appropriate contractor prints, distributes, and assists with scoring the PBT materials. Test materials are produced and shipped to school districts, and secure handling of these materials is required at all times. Secure materials are tracked using unique security numbers assigned to documents, and preidentification (PreID) labels that contain student-level information (e.g., name, FLEID) are affixed to the books prior to testing. The information on these labels ensures that the responses scanned in that book are assigned to the correct student for scoring.

Sample PBT test materials are available on the [FSA Portal](#) to familiarize students, educators, and parents/guardians with the item types and format of each test.

4.3. Computer-Based Testing (CBT)

The following assessments are CBT administrations:

- Grades 7–10 and Retake FSA ELA Writing
- Grades 7–10 and Retake FSA ELA Reading
- Grades 7–8 FSA Mathematics
- FSA Algebra 1 and Geometry EOC Assessments
- NGSSS Biology 1, Civics, and U.S. History EOC Assessments

The appropriate 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. All students taking a CBT test are required to participate in a practice test session prior to testing to familiarize them with the various item types and features of the CBT they will take. Computer-based practice tests are available on the [FSA Portal](#) or on [PearsonAccess Next](#).

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

4.4. Timing

Tests are composed of “sessions,” and each session is conducted in an allotted amount of time. 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 charts below detail the number of sessions, number of days (some two-session tests have both sessions administered on the same day), minutes per session, and total minutes per test.

Table 19: FSA English Language Arts – Writing

Grade Level	Number of Days of Testing	Number of Sessions	Minutes Per Session	Total Minutes Per Grade Level
4	1	1	120	120
5	1	1	120	120
6	1	1	120	120
7	1	1	120	120
8	1	1	120	120
9	1	1	120	120
10	1	1	120	120
Retake*	1	1	120	120

* 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.

Table 20: FSA English Language Arts – Reading

Grade Level	Number of Days of Testing	Number of Sessions	Minutes Per Session	Total Minutes Per Grade Level
3	2	2	80	160
4	2	2	80	160
5	2	2	80	160
6	2	2	85	170
7	2	2	85	170
8	2	2	85	170
9	2	2	90	180
10	2	2	90	180
Retake*	2	2	90	180

* 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.

Table 21: FSA Mathematics

Grade Level	Number of Days of Testing	Number of Sessions	Minutes Per Session	Total Minutes Per Grade Level
3	2	2	80	160
4	2	2	80	160
5	2	2	80	160
6	2	3	60	180
7	2	3	60	180
8	2	3	60	180

Table 22: FSA End-of-Course Assessments

Subject	Number of Days of Testing	Number of Sessions	Minutes Per Session*	Total Minutes Per Assessment*
Algebra 1	2	2	90	180
Geometry	2	2	90	180

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

Table 23: NGSSS Assessments – Science

Subject	Number of Days of Testing	Number of Sessions	Minutes Per Session	Total Minutes Per Assessment
Grade 5 Science	2	2	80	160
Grade 8 Science	1	2	80	160

Table 24: NGSSS Assessments – End-of-Course Assessments

Subject	Number of Days of Testing	Number of Sessions	Minutes Per Assessment*
Biology 1	1	1	160
Civics	1	1	160
U.S. History	1	1	160

* 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.

4.5. 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 contractors.

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.

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. For CBT administrations, a CBT coordinator/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. Test administrators must be employees of the school district and are usually classroom teachers. They must remain in the testing room at all times.

Proctors are recommended at all times, 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 test administrators 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.

Test administrators do not administer tests to their family members. Students related to their assigned test administrator are reassigned to an alternate test administrator. 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, test administrators must sign a *Test Administrator Prohibited Activities Agreement*, which provides a list of actions for test administrators to avoid in order to prevent student test invalidation and/or investigation for misconduct.

4.6. 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 answer books, CBT test tickets, and used work folders, worksheets, or 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 test and answer books and passage booklets, 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. No more than three people should have access to the locked storage room and secure materials must never be left unsecured and must not remain in classrooms or be taken off the school’s campus overnight.

In addition, all content of Florida’s statewide assessments is secure and test items, which result from the detailed, rigorous development process described in previous sections, may be reused on future test forms. Because of this, members of the public may not view tests after the administration 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.

4.7. Administration Procedures

Each PBT and CBT 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 tests:

- [**Planning Sheets**](#) are provided for all students participating in Grades 4–10/Retake ELA Writing (CBT and PBT). Students may use their planning sheets to plan and prewrite their responses.
- [**CBT Worksheets**](#) are provided to all students taking a CBT ELA Reading test to take notes during the test. Students taking PBT tests may make notes in their test and answer books.
- [**Work Folders**](#) are provided to all students taking a CBT Mathematics test or Biology 1 EOC to work the problems during testing. Students taking PBT tests may use the space in their test and answer books.
- [**Scientific calculators**](#) are provided for Grades 7–8 FSA Mathematics (scientific, sessions 2 and 3 only), FSA Algebra 1 and Geometry EOCs (scientific, session 2 only), and

NGSSS Biology 1 EOC (four-function). Calculators are not permitted for Grades 3–6 Mathematics.

Once students are in a testing room, the test administrator will read the administration script verbatim to students. Each test has its own script that 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).
- **PBT Instructions**—Scripts contain directions for completing the front cover and unsealing a test and answer book.
- **CBT Instructions**—Scripts contain directions for logging into a computer-based test, accessing the help menu in the test, and navigating, pausing, and submitting the test.
- **Testing Rules and Testing Rules Acknowledgment**—Students are read the following prior to testing (this example is from Writing; other assessments have minor differences):

During this test, you must not:

- *talk to other students or make any disturbance*
- *look at another student’s test and answer book or planning sheet*
- *allow another student to look at your test and answer book or planning sheet*
- *ask for help writing your response*
- *give help to another student in writing his or her response*
- *have notes or scratch paper other than your planning sheet*
- *have any electronic or recording devices in your possession at any time, including breaks, even if you do not use them*
- *fail to follow any other instructions given*

Because the content in all statewide assessments is secure, you may not discuss or reveal details about the writing prompt or passages after the test. This includes any type of electronic communication, such as texting, emailing, or posting online, for example, on Facebook, Twitter, Snapchat, or Instagram.

Once this portion of the script is read aloud, students are asked to sign below a Testing Rules Acknowledgment (or check a box for some CBTs) that states, *I understand the testing rules that were just read to me. If I do not follow these rules, my test score may be invalidated.*

- **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 [FSA Portal](#) and the [PearsonAccess Next Florida Resource Center](#)

4.8. Accommodations

If test administrators are administering tests to English Language Learners (ELLs) or students with disabilities who have allowable accommodations documented on an Individual Education Plan (IEP) or Section 504 Plan, care must be taken to provide the exact accommodation needed. Appendix A of each test administration manual lists allowable accommodations and related policies. In addition, an [FSA Accommodations Guide](#) is provided with comprehensive information to support successful test administrations to students with disabilities and ELLs.

4.9. 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 for test invalidation are:

- Students are in possession of an electronic device during testing.
- Students are cheating.
- The incorrect amount of time is provided for a test session.
- Students are provided with accommodations that are not allowable or are not listed on the IEP or 504 Plan.
- Students are not provided an accommodation listed on the IEP or 504 Plan.
- Students are given unauthorized assistance during testing.

If a student's test is invalidated, the student may not take the test again in the same administration window. Invalidation codes that may appear on student reports are listed in [Understanding FSA Reports](#) and [Understanding NGSSS Science and EOC Reports](#).

5.0 Scoring the Assessments

The process of scoring statewide assessments begins after the end of each testing window when CBT tests are processed and PBT tests are returned to the contractor. Just as test construction can be viewed in terms of item development and whole-test construction, so can the scoring process 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. In addition, each section details the quality control processes used to ensure the accuracy of scores.

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 establishes the standards in State Board of Education Rule based on these recommendations. 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 two-part items, students must answer both parts correctly to receive full credit for that item.

Gridded-response, open-response, and equation builder 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 to ensure that all possible 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 chose or provided an answer that was not originally identified as correct. These situations lead scoring personnel to investigate whether there is more than one correct answer to a multiple-choice item or whether the list of acceptable answers to gridded-response (PBT) or equation editor (CBT) items may need to be expanded.

Quality Assurance Measures: Statistical Reviews—The same statistical reviews conducted on items after field testing and on test forms during test construction are conducted after operational testing. These reviews are conducted again because the population of students taking the operational test may not have the same characteristics as the field-test population. Another purpose of these reviews is to ensure that the items and test have the characteristics that will make each assessment an effective measure of student achievement. Any deviation from the specified criteria might compromise the accuracy of the student scores.

5.2.2. Handscoring – FSA ELA Writing

Grades 4–10 ELA Writing responses are evaluated through a process called handscoring. Handscoring is guided by a set of Handscoring Specifications. Because the Handscoring Specifications contain detailed information about the ELA test content, they are protected by test security statutes and are not available to the public.

Those selected as professional scorers work in teams of approximately 10 members with each team having a Scoring Supervisor called a Team Leader. Groups of teams score the same grade-level writing prompt at one location. A Scoring Director and an Assistant Scoring Director supervise all the teams assigned to a prompt. Prior to the scoring sessions, all student responses to writing prompts are scanned electronically. At the scoring centers, scorers work individually at computer workstations to read the scanned student responses assigned to them on their computer monitors.

To guide them in rating responses, scorers 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, Focus, and Organization
 - Evidence and Elaboration
 - Conventions of Standard English
- Anchor papers with annotations—Actual, unedited student responses to the task or essay 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 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 require additional training.

All Grade 10 ELA Writing responses are handscored by two human raters, and Grades 4–9 ELA Writing responses are scored by one human rater with 15% of the responses scored by two raters to ensure the validity of the scores.

5.2.3. Reported Scores

All FSA and NGSSS assessments are reported with the following information:

- **Scale scores**—Each assessment has a score scale, and student performance is presented by where the student’s score falls on that scale.
- **Achievement (Performance) Levels**—Scale scores fall into five achievement levels that provide descriptions of the student’s performance.
- **Reporting Category Subscores (raw scores)**—Each reporting category represents groups of similar skills, or standards, that are assessed within each grade and subject. Reporting category performance is conveyed by displaying the points earned and the points possible for each category.
 - For Grades 4–10 and Retake ELA only, the number of points earned in each Writing domain are also reported.

The lowest scale score in Achievement Level 3 is considered the “passing” score for each assessment; however, the only state-level passing requirements apply to the Grade 10 ELA and Algebra 1 EOC Assessments. Passing these assessments is required for graduation with a standard high school diploma. Students who participated in the first administration of one of these 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. Table 25 and Table 26 provide details for Grade 10 ELA and Algebra 1 EOC passing requirements.

Table 25: Grade 10 ELA Passing Score by Year

School Year When Assessment Requirements Began for Students Entering Grade 9	Assessment that Students Must Pass to Graduate	Passing Score for the Required Assessment
2010–2011 to 2012–2013	Grade 10 FCAT 2.0 Reading	245
2013–2014	Grade 10 FSA ELA	349
2014–2015 to current	Grade 10 FSA ELA	350*

* Students who took the assessment prior to the adoption of the passing score on the new scale adopted by the State Board are eligible to use the alternate passing score for graduation, which is linked to the passing score for the previous assessment requirement

Table 26: Algebra 1 EOC Passing Score by Participation Year

First Participation in FSA Algebra 1 EOC	Passing Score
Spring 2016 and beyond	497
Spring, Summer, Fall, or Winter 2015	489*

* Students whose graduation requirement is the NGSSS Algebra 1 EOC may satisfy this requirement by earning the alternate passing score of 489 on the FSA Algebra 1 EOC Assessment.

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](#).

5.2.4. IRT Scoring

Scale scores are the result of a two-step process that analyzes student responses using Item Response Theory (IRT) and uses the resulting item parameter estimates to convert student responses to a scale score that is comparable across test years. As described in IRT Framework on page 31, the IRT model assumes that student responses to items are the result of underlying levels of knowledge and skills, what IRT calls ability and item characteristics. 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. The IRT model is widely used because it produces the most accurate score estimates 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 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 students' overall scale score. Students who correctly answer exactly the same items would, of course, receive the same scale score.

The FSA is comprised of many different item types; some are worth one point (single-point items) and others can be worth more (multi-point items). However, scoring weights of the FSA items are unrelated to how much an item contributes to a student's scale score. Some multiple-choice items may have more weight than some open-ended items, and vice versa. The degree to which an item is weighted is based on the statistical qualities of the item itself.

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.

5.2.5. Process

In the first step of scoring, each item's IRT parameter estimates are calculated using a carefully selected sample of schools that represents the total state population. This is called the calibration sample and the schools selected as part of this sample are often referred to as calibration schools. The role that the calibration schools play is critical to the scoring process because the item parameters that are calculated based upon this sample are used to generate scores for all students.

Equating

After IRT calibration, the process of equating is used to place IRT-processed scores on the assessment scales and to ensure that the resulting scores are comparable to those of previous years. Making scores comparable allows comparisons between, for example, the achievement of grade 6 students in 2018 and the achievement of grade 6 students in 2019. Each assessment is designed to be of similar difficulty each year; however, slight differences in test difficulty (the content of the test items) may influence student scores. Without equating, it would be difficult to determine whether differences in scores between years are the result of these slight differences in the test difficulty or differences in students' true levels of knowledge and skill.

Test developers can isolate the influence of differences in student ability through the use of anchor items—items that appear in tests of consecutive years. Because these items are identical, differences in achievement between groups can be more clearly identified. Using the Stocking/Lord procedure (Stocking & Lord, 1983)⁴, which is the procedure used to maintain the score scales year after year, a statistical relationship is established between the performance of current year students on these anchor items and the performance of students in previous years. This relationship enables the item parameters from the current test form to be expressed on the same scale as the first operational test form. Numerous steps are taken to ensure that the anchor items sufficiently represent the tests so that this relationship can be applied to the entire test for current year students. After this equating process, it is possible to report scores that are comparable to scores of previous years. This means that any differences in scores, such as the difference between mean scores for any two years, can be attributed to differences in student achievement and not to differences in the test difficulty. Anchor items are not included as part of a student's score; they are used only for the purpose of equating. For more details about the equating, please refer to the [FSA technical reports](#).

It is important to emphasize that the cross-year comparability of scores does not extend to the reporting category subscores. These subscores are simply the total of all raw score points

⁴ Stocking, M. L. & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Measurement*, 7, 201–210.

awarded in a given reporting category. Although anchor items are designed to be representative of the test overall, they are not sufficient for making comparisons across years within reporting categories. Such a comparison would require a greater number of anchor items.

Achievement Level Classifications

Based on their scale scores, students are assigned 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 by a special committee, the Standards Setting Committee comprised of Florida educators, as well as FDOE staff, the Florida Education Commissioner, and the State Board of Education. 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 is 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 FSA ELA and FSA 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.

Data Forensics

FDOE employs a data forensics company to review the results for testing irregularities and anomalous data. Possible examples of testing irregularities include a student copying another student's answers or a Test Administrator changing students' answers. Anomalous scores would have included students with similar response patterns in the same testing group or an unusual increase in school performance. The data forensic analyses detect potential security breaches using several statistics to detect the following:

- Pairs or groups of extremely similar or even identical answers
- Aberrant response patterns, such as answering difficult items correctly and not providing correct answers for easy questions
- Response time stamps, in the case of computer-based tests, to check whether a pair or a group of students worked in a synchronous manner
- Unusual gain scores

Through the results of these statistical analyses, it is possible to detect the source of suspect activity and its effect on test results. If an irregularity is found in the data, flagged student records are put on hold and FDOE staff review the data. If student results will not be released, they will be assigned the condition code of NR8, Caveon Invalidated. Districts may conduct an investigation and submit an appeal, with required documentation, to FDOE by an established deadline if the school believes that a student score should be released.

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. Parents and educators seeking a thorough understanding of reports should review the publication [Understanding FSA Reports](#) and [Understanding NGSSS Science and EOC Reports](#). These documents are issued each spring and can be found on the FDOE website.

6.1. Reports

Individual Score Reports

The **Individual Score Report** is a four-page color report for FSA ELA and Mathematics and a three-page color report for the FSA EOC and NGSSS assessments. The report provides general information about the FSA/NGSSS program and the student's results, including the student's scale score, performance level, previous performance (for FSA ELA and FSA Mathematics only, if available), and reporting category scores. 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 a list of helpful resources. The information on this report is presented in English, Spanish, and Haitian Creole.

Also provided on the report is a list of information from the FDOE website that parents/guardians and students may find useful to understand more about the FSA program, state policies regarding assessments, and other resources to promote better understanding of student expectations and to promote family engagement.

Hard-copy Individual Score Reports are provided to school districts, and school districts distribute reports to parents. Districts are also provided electronic copies of the reports for their records, to upload to parent portals, or to create duplicate reports, if needed.

Quality Assurance Measures: Score Inquiry System—If a score for a student is missing, not reported, or is thought to be anomalous, that student's school district may submit a score inquiry to the department. The department and the contractor will research the case, and provide the school district with the results. Examples of possible outcomes of inquiries are a score is missing because a test was returned incorrectly and never processed or confirming with ELA specialists that a student writing response received the correct score.

School, District, and State Reports

The **School Report of Students** for each assessment is available in the secure reporting system for each test vendor and may be accessed by school or district staff with login credentials. Reports are produced for all assessments and contain results, listed by grade level, for all students tested within the school. In addition, a district-level School Report of Students is

available for district users, and contains the School Reports of Students for all schools in the district.

The **District Report of Schools**, **District Summary**, **State Report of Districts**, and **State Summary** contain the number of students, the mean scale score, and the percentage in each performance level for each school, each district, and the state, as applicable. School and district staff with login credentials can access these reports in the secure reporting systems.

Each spring, state reports of districts and state reports of schools are posted to the FDOE website. These Excel files provide the same information as in the District Report of Schools and State Report of Districts. The release includes the spring results along with the summer, fall and winter results in that year. It also includes district comparison spreadsheets that show how districts improved or declined compared to the previous school year as well as summary results packets that show trend performance for all students tested, specific grade bands (elementary, middle, high), and student subgroups (race/ethnicity, students with disabilities, ELLs).

Reporting Categories

The department also provides files to districts and the public that contain the **mean points earned by reporting category** for each assessment. Reporting category and domain scores are the actual number of points possible within each reporting category/domain. These scores are also known as raw scores. Reporting categories represent groups of similar student skills, or standards, which are assessed within each course or subject area. The state mean for each reporting category/domain is also provided to allow for comparisons to average performance statewide. To see what these reports look like, you can access [previous files](#) for FSA ELA, Mathematics, and EOCs, Statewide Science Assessments, and NGSSS EOC Assessments.

The department encourages educators to use statewide assessment results in a way that is statistically appropriate. For educators to gather reliable information from the assessment scores, it is important to identify the comparisons at the reporting category score and/or domain level that yield valid interpretations of student performance. One valid comparison is of performance in a given reporting category or writing domain among schools, districts, and the state. A school's reporting category/domain score results can be compared to other schools', districts', or the state's reporting category/domain score results. District results can be compared to other district results and state results. Because students in any group (school, district, or state) will take the same set of test items in a given year, regardless of varying item difficulty at the content-area level, their results are comparable. For EOCs only, the items in each assessment reporting category will potentially vary by test form. Consequently, users should not compare reporting category scores across forms.

At the student level, reporting categories should not be considered the sole indicators for determining the educational needs of students. Furthermore, providing instruction in a specific reporting category may not be justified and may actually be an inefficient use of instructional time.

When interpreting student results provided under the performance details for each reporting category, the following cautions and information should be considered:

- The number of items in a reporting category will vary by grade level. Consequently, users should not compare reporting category scores across grade levels.
- The difficulty of the items measuring each standard will vary from one year to the next. Consequently, users should not compare reporting category scores across years.

6.2. Florida's PK–12 Education Information Portal

Florida's Education Information Portal, commonly referred to as EDStats, connects people with reports and statistical information about statewide assessment results, and results are available for the most recent five years of an assessment. The portal provides access to standard and interactive reports that offer overall and demographic information in a variety of formats including graphs, tables, maps 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. It can be accessed on the FDOE website at <https://edstats.fldoe.org>. Users are encouraged to review the [User Manual](#) to understand the functionality and full capabilities of the information portal.

6.3. Florida School, District, and State Report Cards

School, District and State Report Cards were developed by the Florida Department of Education (FDOE) 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 [EduData Portal](#). 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 a separate entry.

Achievement Levels: Five categories of achievement that represent the success students demonstrate with content assess on the **FSA** and/or **NGSSS** assessments. Achievement Levels are established using the input of classroom teachers, curriculum specialists, education administrators, and other interested citizens. These professionals helped the Department of Education identify the score ranges for each Achievement Level. The Achievement Levels are helpful in interpreting what a student's **scale score** represents.

Anchor Items: Items which have appeared on a given test in previous years and are used to ensure that the scores on that test can be equated or made comparable from year to year.

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.

Bias: Advantage or disadvantage conferred upon groups of students because of certain personal characteristics (such as gender, race, ethnicity, religion, socioeconomic status, disability, or geographic region) 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 Sample: Carefully selected group of students representative of all students statewide whose response data are used to generate **Item Response Theory (IRT)** parameters used in **operational** testing.

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

Cluster (content cluster): A grouping of related **benchmarks** from the **FSA** and **NGSSS**, clusters are currently used to summarize and report achievement for assessments.

Cognitive Complexity: System used to classify **items** according to the complexity of the steps and processes they require students to use.

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**, **writing samples**, 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. The **FSA** and **NGSSS** assessments, both **CRTs**, measure student progress toward meeting these **standards**.

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

Depth of Knowledge: See **Cognitive Complexity**.

Differential Item Functioning: Statistical technique in which gender and ethnic **bias** can also be identified in the statistical analysis of field and operational test data.

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

Editing Task Choice: Items with a highlighted word or phrase, which reveals a drop-down menu containing options for correcting an error as well as the highlighted word or phrase as it is shown in the sentence to indicate that no correction is needed.

Equating: A process used to place **IRT**-processed scores on the assessment scales and to ensure that the resulting scores are comparable to those of previous 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, F.A.C.

Florida Standards Assessments (FSA): Statewide assessments which measure whether students have made progress on the English Language Arts (ELA) Florida Standards and the Mathematics Florida Standards.

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.

FSA Portal: Resources and information for district and school personnel are located on the FSA Portal, which is accessed at www.FSAssessments.org.

Graphic Response Item Display: Item in which a student may select numbers, words, phrases, or images and use the drag-and-drop feature to place them into a graphic organizer or other format.

Gridded-Response Adjudication Committee: Members review all responses to field-tested gridded-response items to determine whether all possible correct answers have been included in the scoring key.

Gridded-Response Item: Test **items** that require students to solve a problem for which the answer is numerical. Answers must be written and gridded into a number grid. The gridded-response item format is used in **FSA Mathematics**.

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**. **Items** are selected from it each year to construct the **FSA** and **NGSSS** assessments.

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): Statistical model for student responses to test **items**. 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 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 correct answers from among the options provided.

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

Next Generation Sunshine State Standards (NGSSS): Assessments which measure student achievement of the **benchmarks** of the NGSSS Science Standards and the NGSSS Social Studies Standards.

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 method of calculating a test score based on the pattern of correct and incorrect answers provided by a student. A student's pattern of answers is analyzed in combination with the **IRT item** parameters.

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 to for **FSA ELA Writing**.

Psychometrics: The field of study 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 **FSA** Grade 10 ELA assessment given to those who do not achieve the passing score required for high school graduation.

Sample Item: A writing sample is an example of draft writing. Writing samples may be draft stories, reports, or articles that contain some mistakes. **FSA Writing items** based on writing samples ask about the strengths and weaknesses of the sample.

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

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 **test and answer books** and test tickets used for computer-based testing.

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.

Strand: The broad divisions of **content areas**. For example, in the **FSA** ELA, there are seven strands: Reading, Writing, Listening, Viewing, Speaking, Language, and Literature.

Table Item: In Mathematics only, an **item** type in which a student types numerical values into a given table.

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 Answer Books: For **FSA** paper-based tests, students are provided test and answer books that contain all of the **items**. Students respond directly in their books.

Test Construction: Process of selecting **items** to build test forms for spring test administrations.

Test Construction Specifications: Specifications based on test **item** specifications and other considerations such as statistical criteria. The test construction specifications are revised annually to guide the construction of each assessment.

Test Form: A unique set of **items** consisting of core operational **items** and a smaller number of either **field-test** or **anchor items**. Students with different test forms encounter the same **operational items**, but different **field-test** or **anchor items**.

True Score: The **FSA** and **NGSSS** assessments seek to measure a student’s “true” achievement or true score on the content assessed. By definition, a student’s test score is composed of

two parts: the true score and the **standard error of measurement** associated with the test.

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.

Writing Samples: Writing samples may be draft stories, reports, or articles that contain some mistakes. **FSA Writing items** based on writing samples ask about the strengths and weaknesses of the sample.

8.0 Guide to Useful Resources

8.1. ELA Writing Resources

Several interpretive resources are available to provide examples and information on how FSA ELA 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.

- [FSA Fact Sheets](#)
- [Statewide Science Fact Sheet](#)
- [NGSSS EOC 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. Practice Tests

The purpose of the practice tests 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 practice tests are not intended to guide classroom instruction.

- [FSA CBT Practice Tests](#) (and guide)
- [FSA PBT Practice Tests](#)
- [Statewide Science Sample Test Books and Answer Keys](#)
- [NGSSS EOC CBT Practice Tests](#)
- [NGSSS EOC PBT Sample Tests](#)

8.5. Technical Reports

The statewide assessment 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.

- [FSA Technical Reports](#)
- NGSSS Technical Reports

8.6. Test Administration Manuals

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

- [FSA Test Administration Manuals](#)
- [NGSSS Test Administration Manuals](#)

In addition to the manuals, an [FSA 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

Test Design Summaries 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.

- [FSA Test Design Summaries](#)
- [NGSSS Test Design Summary](#)

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.

- [FSA Test Item Specifications](#)
- [Statewide Science Test Item Specifications](#)
- [NGSSS EOC Test Item Specifications](#)

8.9. Schedules

[Statewide Assessment Schedules](#) are established according to Florida Statute 1008.22 for FSA and NGSSS assessments. These schedules also contain testing dates and windows for other statewide assessments, such as FSAA, ACCESS for ELLs, and NAEP.

8.10. Standard Setting

As assessments are implemented for the first time, a standard setting process is required. 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. 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. Aggregate Assessment Results

FSA and NGSSS assessment results are provided in various formats to accommodate the needs of parents, educators, districts, and researchers. Historical results are posted by year on the [FDOE website](#), and the most recent five years of results are provided in an interactive reporting database, commonly referred to as [EDStats](#), which allows users to customize and extract data in a variety of formats.

An additional resource available from FDOE is the [Florida Report Cards](#) (EduData site). The EduData Portal is a resource to help parents and other stakeholders understand how the state, district and schools are doing with regard to student achievement and success specified by the Elementary and Secondary Education Act of 1965 as amended by the Every Student Succeeds Act (ESSA).

8.12. 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 FSA Reports](#)
- [Understanding NGSSS Reports](#)

8.13. Useful Links

- The [FSA Portal](#) is the website containing resources, links, and information for Florida Standards 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.
- [Florida Resource Center](#) is the website containing resources, links, and information about the NGSSS assessments.
- The [Florida Standards Alternate Assessment \(FSAA\)](#) 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\(2\)](#), Florida Statutes.

Appendix A: Statistical Indicators Used in Data Analysis

After field testing, during the test construction process, and after operational testing, a series of statistical analyses is performed on test items and the test as a whole to ensure that established criteria for items and test forms have been or will be met. The purpose of the review is to determine whether individual items can be used in the future as operational items. During test construction, data are reviewed for individual items and proposed test forms. After operational testing, data are generated from a sample of students representative of all students tested (the calibration sample) to generate the parameters necessary for scoring (IRT processing) and to determine whether any items require special treatment in the scoring process. Additional measures are generated after 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 established criteria 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 and test construction 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 FSA and NGSSS 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 gender (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)⁵ for single-point items and the Generalized Mantel-Haenszel chi-square statistic (Somes, 1986)⁶ 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 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

⁵ 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.

⁶ Somes, G. W. (1986). The generalized Mantel Haenszel statistic. *The American Statistician*, 40:106–108.

particular gender 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)⁷, 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 the total score for all items (raw score). 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 (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

⁷ 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.

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 27: 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 Assessment 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 Assessment 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.
- Nation’s first graduation test was authorized in 1976, and implemented first with the graduating class of 1983 (minimum competency test).

1990s

- 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 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 of 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 EOCs 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 measure 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.

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

Table 28 provides an overview of the Florida Statutes (F.S.) and State Board of Education rules (Florida Administrative Code, F.A.C.) that govern Florida's statewide assessment 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 29.

Table 28: Statewide, Standardized Assessments Statutes and Rules

Assessment	Assessment Citation	Required Use	Required Use Citation
Statewide Assessment Program (FSA, EOCs, FSAA, Concordant Scores)	s. 1008.22, F.S. Rule 6A-1.09422, F.A.C. Rule 6A-1.09430, F.A.C. Rule 6A-1.0943, F.A.C. Rule 6A-1.09432, F.A.C. Rule 6A-1.094223, F.A.C.	Third Grade Retention; Student Progression; Remedial Instruction; Reporting Requirements	s. 1008.25, F.S. Rule 6A-1.094221, F.A.C.
		High School Standard Diploma	s. 1003.4282, F.S.
		EOC Assessments as 30% of Course Grade	s. 1003.4282, F.S. s. 1008.22, F.S.
		School Grades	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.
		District Grades	s. 1008.34, F.S.
		Differentiated Accountability	s. 1008.33, F.S. Rule 6A-1.099811, F.A.C.
		Opportunity Scholarship	s. 1002.38, F.S.
		Uniform Assessment Calendar	Rule 6A-1.094224
		Concordant and Comparative Scores for Graduation	Rule 6A-1.09422
		ESE Graduation Requirements and Waiver of Results	Rule 6A-1.09963

Assessment	Assessment Citation	Required Use	Required Use Citation
ACCESS for ELLs, Kindergarten ACCESS for ELLs, Alternate ACCESS for ELLs – English language acquisition assessments to determine ESOL program placement or exit eligibility	<u>s. 1003.56, F.S.</u>	English for Speakers of Other Languages (ESOL) Exit	<u>Rule 6A-6.0902, F.A.C.</u> <u>Rule 6A-6.0903, F.A.C.</u> <u>Rule 6A-6.09021, F.A.C.</u>
Postsecondary Education Readiness Test (PERT) – a common placement test for determining readiness for college instruction in Florida	<u>s. 1008.30, F.S.</u>	College Readiness Coursework	<u>Rule 6A-10.0315, F.A.C.</u>
Preliminary SAT (PSAT)/Preliminary ACT (PLAN) – administered by each public high school to all 10th-grade students, though the parent has the opportunity to exempt his/her child from PSAT/PLAN	<u>s. 1007.35, F.S.</u>	Inform Course Placement	
Kindergarten Screening – required for all kindergarten students within the first 30 days of the school year to determine the readiness and performance of VPK providers	<u>s. 1002.69, F.S.</u>	VPK Readiness Rates	<u>Rule 6M-8.601, F.A.C.</u>
National Assessment of Educational Progress (NAEP) – administered to a sample of students in selected grade levels every other year	<u>s. 1008.22, F.S.</u>	National and State Comparisons	

Assessment	Assessment Citation	Required Use	Required Use Citation
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	

Table 29: 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 Rule 6A-6.0903
Test Security	s. 1008.24, F.S.	Rule 6A-10.042

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 English Language Arts Assessment to be promoted to grade 4.
 - Good cause exemptions are provided by law.
- A student must pass the Grade 10 English Language Arts 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 assessment 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:
 - English Language Arts 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.
- Requires five achievement levels for all statewide assessments, with level 1 being the lowest and level 5 being the highest.
- Prohibits districts from suspending instruction for the sole purpose of practicing administration of statewide tests and lists allowable practice activities.
- Sets requirements for testing schedules, earliest test administration dates, and required reporting dates.
- Includes provision for the Florida Standards Alternate Assessment (FSAA) 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.
- Sets requirements for local assessments (this topic is addressed in this report in a separate section on local assessments).

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 © 2019
State of Florida
Department of State