

Benchmarks for Excellent Student Thinking (B.E.S.T.) Assessments Standard Setting Educator Panel

Grades 3–10 ELA Reading, Grades 3–8 Mathematics, and B.E.S.T. EOC Assessments

Orlando, FL | July 2023



www.FLDOE.org



Presenters

- Welcoming Comments and Setting the Stage for Standard Setting
 - Vince Verges, Assistant Deputy Commissioner, Division of Accountability, Research, and Measurement, FDOE

Standard Setting

 Gary W. Phillips, Vice President for Psychometrics, Cambium Assessment





Topics

- Information about the assessments aligned to the B.E.S.T. Standards
- Review the Standard Setting process
- Review Achievement Level Descriptions (ALDs) & Just Barelys
- Discuss the Bookmark Method of Standard Setting
- Define and discuss articulation
- Define and discuss impact data
- Define and discuss benchmark data



Assessments Aligned to the B.E.S.T. Standards



www.FLDOE.org



Assessments Aligned to the B.E.S.T. Standards

- Assessments in English Language Arts (ELA) and Mathematics measures student achievement of the B.E.S.T. Standards.
 - Information about the B.E.S.T. Standards may be accessed at http://www.cpalms.org.
- The assessments aligned to the B.E.S.T. Standards were first administered during the 2022–2023 school year (baseline administration).





Assessments Aligned to the B.E.S.T. Standards (cont.)

- The following assessments are aligned to the B.E.S.T Standards:
 - Florida Assessment of Student Thinking (FAST) Progress Monitoring Assessments:
 - Kindergarten through grade 10 ELA Reading
 - Kindergarten through grade 8 Mathematics
 - Grades 4 through 10 ELA Writing
 - End-of-Course (EOC) Assessments:
 - Algebra 1
 - Geometry



Administration Information: FAST

- Aligned to the B.E.S.T. Standards and administered as computer-adaptive tests (CAT)
- FAST progress monitoring assessments are administered three times per year:
 - PM1 occurs in the beginning of the school year and serves as a baseline
 - PM2 occurs mid-year and produces a mid-year score to compare against the baseline
 - PM3 occurs at the end of the school year and produces a summative scores
- PM1 and PM2 are used for informational purposes only; only PM3 is used for school accountability beginning in the 2023–2024 school year.



Administration Information: FAST (cont.)

- Assessments are administered in one session on one day.
 - Recommended that students take only one subject test per day
- Paper-based accommodations are offered for eligible students
- Accommodations may vary depending on the subject, method of administration (i.e., computer-based or paperbased), or the progress monitoring session
- Students taking computer-based tests had opportunities to access practice tests to become familiar with item types, functionality, and test mode or platform.





Administration Information: FAST K–2

 Kindergarten through grade 2 FAST assessment are administered using Renaissance Learning's Star Early Literacy, Star Reading, and Star Mathematics.

Grade	Subject	PM1	PM2	PM3
Kindergarten	Reading	Star Early Literacy	Star Early Literacy or Star Reading*	Star Early Literacy or Star Reading*
	Math	Star Math	Star Math	Star Math
Grade 1	Reading	Star Early Literacy	Star Early Literacy or Star Reading*	Star Reading**
	Math	Star Math	Star Math	Star Math
Grade 2	Reading	Star Reading*	Star Reading*	Star Reading*
	Math	Star Math	Star Math	Star Math
*All Kindergarten and Grade 1 students will take the Star Early Literacy for PM1. If a student scores at or above 852 on Star Early Literacy, the student will take Star Reading on subsequent administrations. If a student scores below 852 on Star Reading, the student may take Star Early Literacy in the subsequent administration. **All Grade 1 students must attempt Star Reading in PM3 if not already taking Star Reading.				



Administration Information: FAST

K–2 Assessment		Item Time Limit	Items Per Test	Average Test Length
Star Early Literacy	9	0 seconds per item	27	10–20 minutes
Star Reading	60 seconds per item (first 10 items) 120 second per item (remaining skill items)		34	15–20 minutes
Star Math	3 minutes per item		34	20–30 minutes
Grade		Session		
		PM1/2	PM3*	Max Number Items
3–10 ELA Reading		Up to 90 minutes	Up to 120 minutes	40 items
3–5 Mathematics		Up to 80 minutes	Up to 100 minutes	35 items
6–8 Mathematics		Up to 100 minutes	Up to 120 minutes	35 items

*PM3 is slightly longer than PM1/2 because the assessments contain embedded field test items.



Administration Information: B.E.S.T. ELA Writing

- Aligned to B.E.S.T. Standards
- First administered in Spring 2023 as a standalone field test administered to a representative sample of Florida students
- Beginning with the 2023–2024 school year, ELA Writing will be administered during each spring administration.
- ELA Writing scores will be reported separately from ELA Reading; they will not contribute to an overall ELA score.





Administration Information: B.E.S.T. EOC Assessments

- Includes Algebra 1 and Geometry
- Aligned to the B.E.S.T. Standards
- B.E.S.T. EOC assessments are not part of the FAST progress monitoring program.
- First administered as computer-adaptive test (CAT) beginning Winter 2022





Administration Information: Writing and EOCs

Grade/Subject	Session Length	Max Number Items
Grades 4–10 ELA Writing	Up to 120 minutes	N/A
Algebra 1 EOC	Up to 160 minutes	45 operational items
Geometry EOC	Up to 160 minutes	45 operational items





The Standard Setting Process



www.FLDOE.org



Why Are Standards Necessary?

- To define what students should know and be able to do
- To identify clear expectations for students, parents, and teachers
- To improve teaching and learning
- To develop a society able to compete in a global economy





Types of Standards

- <u>Content Standards: Define desired student knowledge</u> and skills (the "what")
 - B.E.S.T. Standards
- <u>Achievement</u>: Describe how much content knowledge a student is required to demonstrate
 - Achievement Level Standards
 - Assessment Graduation Requirements (Algebra 1 and Grade 10 ELA)
- <u>Accountability Standards</u>
 - School Grading Criteria
 - Annual Measurable Objectives



What is Standard Setting?

- A process of deriving levels of performance on educational or professional assessments, by which decisions or classifications of persons will be made (Cizek, 2006)
- Test scores can be used to group students into meaningful Achievement Levels.
- Standard setting is the process whereby we "draw the lines" that separate the test scores into various Achievement Levels.
- A process that is required when implementing new standards and new assessments.





Setting Standards Is Aspirational

- Standard setting is all about what students should know and be able to do, not about what they currently know and are able to do.
- The goal is to set standards for **all applicable students** across the state.





Achievement Levels

• There are five Achievement Levels, which requires the setting of four Achievement Level cuts.

Five Achievement Levels, Four Cut Points



Level 3 indicates On Grade Level performance on each assessment.





Standard Setting: A Multi-Stage Process





We've Done This Before...

1998

- Reading and Mathematics Achievement Levels approved for grades 4, 5, 8 and 10
 2001:
- Reading and Mathematics Achievement Levels approved for grades 3–10
- Grade 10 FCAT Reading and Mathematics passing scores established
 2011:
- FCAT 2.0 Reading (grades 3–10) and Mathematics (3–8) Achievement Levels approved
- Algebra 1 EOC Assessment Achievement Levels approved
- Grade 10 FCAT 2.0 Reading and EOC assessment passing scores established in rule as the minimum score in Achievement Level 3

2012:

- FCAT 2.0 Science (grades 5 and 8) Achievement Levels approved
- Biology 1 and Geometry EOC assessments Achievement Levels approved
 2013:
- U.S. History EOC Assessment Achievement Levels approved **2014**:
- Civics EOC Assessment Achievement Levels approved
 2016:
- Florida Standards Assessments (FSA) ELA, Mathematics, and EOC (Algebra 1, Geometry, Algebra 2) Achievement Levels approved





Standard Setting Terminology

- Academic Standards: The content that students are expected to know.
- Achievement Levels: Levels of student achievement based on observed scale scores.
- Achievement Level Descriptions (ALDs): Descriptions of the competencies associated with each level of achievement.
- Cut Scores (Standards): Scores on an assessment that separate one Achievement Level from another.
- Panelists (Judges/Raters): Those who participate in the standard setting process (stakeholders, educators, professionals).





Standard Setting Terminology (cont.)

- Feedback Data: Data provided to panelists to help them assess the validity and reasonableness of the standards they are recommending (e.g., median/mean cut score ratings, degree of panelist agreement).
- Impact Data (Normative Feedback): Data that summarize the consequences of a proposed set of cut scores (e.g., How many students' scores will be classified at Level 3?).
- Benchmark Data: Data that summarize Florida students' performance on national and international standardized assessments; helps determine whether achievement standards are nationally and internationally competitive.





Important Dates

- Achievement Level Descriptor (ALD) Panels
 - April & October 2021
 - Virtual
- Educator Panel
 - July 24–28, 2023
 - Orlando, Florida
- Reactor Panel
 - August 3–4, 2023
 - Orlando, Florida



ALD Panels

- ELA Reading and Mathematics ALD Panels: K–2
 - March 2023: K–2 ELA Reading and Mathematics
 - 6 panelists for each subject K–2
- ELA Reading and Mathematics ALD Panels: 3–10 & EOCs
 - April 2021: Grades 3–10 ELA Reading
 - October 2021: Grades 3–8 Mathematics, Algebra 1 and Geometry EOC Assessments
 - Approximately 30 panelists



Educator Panels

- Grades 3–10 ELA Reading, Grades 3–8 Mathematics, B.E.S.T. EOC Assessments
 - July 24–28, 2023; Orlando, Florida
 - Five-day Standard Setting workshop
 - Four rounds of standard setting
 - Approximately 320 panelists
 - 16 rooms setting standards concurrently
- Kindergarten–Grade 2 ELA Reading and Mathematics
 - July 25–27, 2023; Orlando, Florida
 - Moderation with table leaders: July 28; Orlando, Florida
 - Three-day Standard Setting workshop
 - Three rounds of standard setting
 - Approximately 120 panelists
 - 6 rooms setting standards concurrently



Educator Panels (cont.)

- Grades 4–10 ELA Writing Assessments
 - July 25–27, 2023; Orlando, Florida
 - Three-day Standard Setting workshop
 - Found rounds of standard setting
 - Approximately 140 panelists
 - 7 rooms setting standards concurrently
 - 1 cut score (On Grade Level), 5 achievement levels



Reactor Panel

- August 3–4, 2023; Orlando, Florida
- Two-day meeting composed of
 - Community/education organization leaders
 - State university leaders
 - Business leaders
 - School board members
 - Superintendents
- Approximately 10–20 panelists
- Review educator panel recommended cut scores and impact data
- Two rounds of review



Main Activities of the Educator Panel

- Table leader training
- Large group orientation
- Grade/subject specific training
- Panelists will
 - take the online test in assigned subjects/grades,
 - review the content standards,
 - review achievement level descriptions,
 - created "just barely" summary ALDs, and
 - review the ordered item booklet.
- Recommend four achievement standards in four rounds
 - Level 2, Level 3, Level 4, and Level 5
- Workshop evaluation



Achievement Level Descriptors and Just Barelys



www.FLDOE.org



Achievement Level Descriptors

- ALDs specify what students in each achievement level are expected to know and be able to do.
- ALDs are the link between content and achievement standards.
- Use the ALDs to develop a mental representation of students at each achievement level.
- Place the bookmark at the point where students scoring at and above that level can be accurately described by the ALD.





Achievement Standards

Level	Label
Level 5	Exemplary
Level 4	Proficient
Level 3	On Grade Level
Level 2	Approaching Grade Level
Level 1	Emerging





Achievement Standards (cont.)





Other Types of Achievement Level Descriptors

- Policy achievement level descriptors Statements about the state's vision and intended policy goals for the achievement levels. In Florida, the Policy ALDs are referred to as Achievement Level Policy Definitions.
- **Reporting achievement level descriptors** Brief summaries of what students know and can do in each achievement level. In general, the Reporting ALDs are summaries of the salient features of the Range ALDs.





	Policy	Achievement	Level	Descriptors
--	--------	-------------	-------	-------------

Emerging (Level 1)	Students at this level demonstrate emerging skills but are not yet demonstrating On Grade Level success with the challenging content on the <i>B.E.S.T. Standards</i> .	
Approaching Grade Level (Level 2)	Students at this level demonstrate that they are approaching but are not yet demonstrating On Grade Level success with the challenging content of the <i>B.E.S.T.</i> <i>Standards</i> .	
On Grade Level (Level 3)	Students at this level demonstrate On Grade Level success with the challenging content of the <i>B.E.S.T. Standards</i> .	
Proficient (Level 4)	Students at this level demonstrate a proficient level of success with the challenging content of the <i>B.E.S.T. Standards</i> .	
Exemplary (Level 5)	Students at this level demonstrate exemplary success with the most challenging content of the <i>B.E.S.T. Standards</i> .	
<u>a</u>	www.FLDOE.org	35



Just Barely ALDs

- When considering each achievement level, we are interested in those students who just barely reach the standard
 - Not typical of students in achievement level. Although just barely, they do reach the standard.




The Bookmark Method





Standard Setting Method

- Bookmark Method One of a number of approaches available for standard setting
 - One of the most widely used and most defensible methods
 - Research-based procedure
 - Used in many state assessment programs
 - Proven to be technically sound in litigation

Mitzel, H. C., Lewis, D. M., Patz, R. J. & Green, D. R. (2001). The Bookmark procedure: Psychological perspectives. In: G. Cizek (Ed.), Setting performance standards: Concepts, methods and perspectives. Mahwah, NJ: Erlbaum.





Bookmarking Pages in the Ordered Item Booklet

- The terms "bookmarking pages" and "ordered item booklet" (OIB) historically come from a pencil/paper testing environment.
- For our standard setting, the ordered item booklet is online so the pages selected will be from a drop-down menu.
- As you study the OIB consider each item and answer two questions:
 - What do students need to know and be able to do to respond successfully to this item?
 - Why is this item more difficult than the previous items?



Bookmark Method Using the OIB

- OIBs are used in the Bookmark Method to provide a content basis for recommending cut scores.
- OIBs will match the test blueprint (for example, if the test is 20% Algebra, then the OIB is 20% Algebra).
- Items in the OIB are ordered by item difficulty .
- Each item, or a point from an item, is presented on a separate page.
- Each item in the OIB is associated with a scaled score on the test. So, when you pick an item for the cut score, you are picking the scaled score in the background that will be the cut score on the test.
- Constructed-response items or multi-point items appear multiple times in the OIB (once for each score point).
- The number of pages in the book is equal to the number of points in the ordered item booklet (not the number of items).





Bookmarking Pages in the OIB







Bookmarking Pages in the OIB (cont.)



C



Chance of Getting an Item Correct

- Imagine that you have 100 students who are "just barely" at or above the On Grade Level cut score.
- For each page in the OIB, you should ask yourself the question, "Out of 100 students who are 'just barely' at On Grade Level, what percent would likely get this item correct or earn this point on a multi-point item?"
 - Toward the beginning of the OIB, you would expect almost all of the students who are "just barely" On Grade Level would likely get the item correct because the items are easy for the On Grade Level student.
 - Toward the end of the OIB, you would expect very few of the "just barely" On Grade Level students would likely get the item correct because the items are difficult for the On Grade Level student.



Chance of Getting an Item Correct (cont.)

- What the concept shows is that a student who is "just barely" On Grade Level will not have an equal likelihood of getting each item correct in the OIB.
- In the Bookmark procedure, the panelist is trying to locate the item, or small group of items, where out of 100 students who are just barely at the On Grade Level, about 50% of them would correctly get this item correct or earn this point on a multi-point item.



Making Your Bookmark

- As you work through the OIB, you will come across an item, or small group of items, where you think about 50% of the "just barely" On Grade Level students would likely get that item correct.
- Look past the selected item to make sure the selected item is the best item.
- Items before that point in the OIB are items that you feel more than 50% of just barely On Grade Level students would get correct.
- Items beyond that point in the OIB are items that you feel less than 50% of just barely On Grade Level students would get correct.
- For each page, ask yourself the question: would 50% of just barely On Grade Level students get this item correct or earn a point on a constructed-response item?
- Panelists place a bookmark on the last page in the OIB where they believe about 50% of the "just barely" students would correctly answer the item or earn a point on a constructed-response item
- This process is repeated for each achievement level.





Making Your Bookmark (cont.)

- For each of the four rounds of standard setting make your recommendations in this order
 - Level 3, On Grade Level
 - Level 2, Approaching Grade Level
 - Level 4, Proficient
 - Level 5, Exemplary



Preparations for Setting Marks

- Practice Test
 - Panelists will take the test exactly as the students took the test using an online tool.
 - The panelists in each room will take the test that is associated with that room. It is possible that some panelists may see items from the testing experience in the Ordered Item Booklet (OIB); however, it's important to note that some overlap of items is expected.
- OIB Practice Round
 - Panelists will use a six-item OIB designed to give them an understanding of how the bookmarking works and how to recommend a cut score. The six items will cover a range of item types and is used for discussion purposes after the panelists complete the practice task.





Room Median

- The recommended standard in each round of standard setting will be the median page number based on all panelists in the room.
- If the median is a fraction, it will be rounded to the nearest whole number.



Mechanics of the Bookmark Method

- Day 1 Training and "Just Barely" ALDs development
- Day 2 Round 1: Content (initial judgment based solely on OIB)
- Day 3 Round 2: Panelists receive feedback from Round 1 and articulation graphs are introduced
- Day 4 Round 3: Impact data introduced
- Day 5 Round 4: Benchmark data is introduced





What if an items seems wrong or unfair?

- All items have been vetted and approved through content and bias reviews.
- The educator panel is not another item-review meeting.
- If a panelist believes something is wrong with an item, they can tell the facilitator or an assistant facilitator; then, they can skip the item to review the rest of the OIB.



What if an item seems out of order?

- Item order is based on student performance.
- Items may seem out of order because they are ordered by difficulty, not by content or cognitive process.
- The sequence of items in the OIB will not match the sequence of instruction taught throughout the school year.
- The ordering of items in the OIB will not match the ordering of items on the test.





Articulation





Vertical Scale versus Within-Grade Scales

- Vertical scales
 - ELA Reading scale, kindergarten through grade 10 (K–2 + 3–10)
 - Mathematics scale, kindergarten through grade 8 (K–2 + 3–8)
- Within-grade scales
 - Algebra 1
 - Geometry



Articulation

- Even though each room will be recommending standards for one grade and one subject, the entire workshop will be recommending a system of achievement standards.
- Achievement standards for a statewide system must be coherent across grades and subjects:
 - Vertically articulated across grades with no anomalous peaks and valleys
 - Orderly across subjects with no dramatic differences in expectation





Articulation (cont.)

- Cambium will fit a regression model to the data and produce a graph representing articulated standards. The regression line may be augmented by the benchmark data.
- The articulated standards will be communicated to the panelists at the beginning of each round after round 1.
- With articulated cut scores in hand, panelists will be asked to judge whether it makes sense from a content point of view to place their bookmark on, or near, the OIB page associated with each articulated cut score.





Illustration of Disarticulated Standards

ELA Reading - Percentage of Students at and Above Each Achievement Standard - Round 4





Illustration of Articulated Standards

ELA Reading - Percentage of Students at and Above Each Achievement Standard - Round 4





Illustration of Disarticulated Standards

ELA Reading - Cut Score in the Scaled Score Metric - Round 4





Illustration of Articulated Standards

ELA Reading - Cut Score in the Scaled Score Metric - Benchmark Legacy Standards





Impact Data and Benchmark Data





Impact Data

- Impact data are the percent of students at and above each selected cut score.
- Data are from students who participated in the Spring 2023 test administration.
- Will be given to panelists in Round 3 after the panelists made judgments across two rounds.
- Impact data will be used as context to inform the panelists' recommendations but will not determine their recommendations.
- In the end, the panelists' recommendations should have a content justification.



Benchmark Data

- Benchmarking provides the panelists with external references so they can see how their recommendations stack up with important external standards.
- Introduced in round 4 for ELA Reading grades 3–10, Mathematics grades 3–8, and Algebra 1 and Geometry EOC assessments
 - Data from the legacy test
 - National Assessment of Educational Progress (NAEP) using Florida results from 2022





Meeting Logistics





Confidentiality

- DO NOT
 - Discuss the test items outside of this meeting
 - Remove any secure materials from the room on breaks or at the end of the day
 - Discuss judgments or cut scores (yours or others) with anyone outside of the meeting
 - Discuss secure materials with non-participants
 - Use cell phones in the meeting rooms (Please turn your cell phone ringer off.)
- "What happens in the meeting room stays in the meeting room."
- General conversations about the process and days' events are acceptable.
- Participants should avoid discussing details, particularly those involving items, cut scores, and any other confidential information.
- Notes should be taken using provided materials only.
- The only materials allowed on the table are standard setting materials.



Educator Panel Day 1 – Monday

C

d

	6
5:00 PM	Adjourn
4:45–5:00 PM	Tables work to produce "Just Barely" statements Panelists fill out daily evaluation
3:30–4:45 PM	Work within tables to produce "Just Barely" statements (<i>submit final document to facilitator</i>)
3:15–3:30 PM	Break
2:00-3:15 PM	Review "Just Barely" descriptors
1:00-2:00 PM	Review Range Achievement Level Descriptors
12:00-1:00 PM	Lunch
11:15–12:00 PM	Panelists experience online operational test environment
11:00–11:15 AM	Break into individual rooms
10:00–11:00 AM	Welcome from FDOE and large group introductory training
8:30–10:00 AM	Table Leader Training and Educator Panel Registration
7:30–8:30 AM	Breakfast and registration



Educator Panel Day 2 – Tuesday

Ś

	7:30–8:30 AM	Breakfast and sign-in	
	8:30–10:00 AM	Wrap up "Just Barely" discussion	
	10:00–10:15 AM	Break	
	10:15-12:00 PM	Review of OIB and training on Standard Setting tool	
	12:00-1:00 PM	Lunch	
	1:00-2:00 PM	Review of OIB (continued)	
	2:00-2:30 PM	Training on Bookmark Placement task	
	2:30-3:00 PM	Training on CAI Standard Setting software	
	3:00-3:15 PM	Break	
	3:15-4:45 PM	Round 1 bookmark placement	
	4:45–5:00 PM	Panelists fill out daily evaluation	
	5:00 PM	Adjourn	
Γ	3	www.FLDOE.org	66
	•		



Educator Panel Day 3 – Wednesday

7:30-8:30 AM	Breakfast and sign-in
8:30–10:30 AM	Feedback and discussion from Round 1
10:30–10:45 AM	Break
10:45–1:00 PM	Round 2 bookmark placement (articulation introduced)
1:00-2:00 PM	Lunch
2:00-4:00 PM	Feedback and discussion from Round 2
4:00-4:15 PM	Break
4:15-4:45 PM	Review of plan for Round 3 and Round 4
4:45-5:00 PM	Panelists fill out daily evaluation
5:00 PM	Adjourn





Educator Panel Day 4 – Thursday

7:30–8:30 AM	Breakfast and sign-in	
8:30–8:45 AM	Feedback and discussion from Round 2	
8:45–10:45 AM	Round 3 bookmark placement (impact data presented)	
10:45–11:00 AM	Panelist break	
11:00-12:00 PM	Feedback and discussion from Round 3	
12:00-1:00 PM	Lunch	
1:00-3:00 PM	Large group discussion	
3:00-3:15 PM	Panelist break	
3:15–4:45 PM	Individual room discussion	
4:45–5:00 PM	Panelists fill out daily evaluation	
5:00 PM	Adjourn	
	www.FLDOE.org	68



Educator Panel Day 5 – Friday

7:30–8:30 AM	Breakfast and sign-in
8:30–10:30 AM	Round 4 bookmark placement (benchmark data presented)
10:30–10:45 AM	Panelist break
10:45–11:45 AM	Feedback and discussion from Round 4
11:45–12:00 PM	Panelists fill out daily evaluation and comprehensive evaluation
12:00-1:00 PM	Lunch
1:00 PM	Adjourn
1:00-3:00 PM	Table leader moderation session (if needed)





Questions?





