



Florida Standards Assessments

2014–2015

Volume 3 Setting Achievement Standards



FLORIDA DEPARTMENT OF
EDUCATION
fldoe.org

ACKNOWLEDGEMENTS

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

Major contributors to this technical report include the following staff from AIR: Dr. Harold Doran, Dr. Elizabeth Ayers-Wright, Dr. Dipendra Subedi, Dr. MinJeong Shin, Dr. AhYoung Shin, Danielle Peterson, and Patrick Kozak. The major contributors from the Florida Department of Education are as follows: Dr. Salih Binici, Dr. Molly Hand, Dr. Qian Liu, Vince Verges, Victoria Ash, Susie Lee, Mengyao Cui, Steve Ash, Renn Edenfield, and Chris Harvey.

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	1
2. BACKGROUND	14
2.1 Achievement Levels	14
2.2 Achievement Level Descriptions	15
3. EDUCATOR PANEL	16
3.1 Educator Panel Composition	16
3.2 Logistical Preparation	31
3.3 Training	33
3.3.1 Taking the Test	33
3.3.2 “Just Barely” Achievement Level Descriptions	33
3.3.3 Ordered Item Booklet	35
3.4 Marking the Pages in the Ordered Item Booklet	37
3.4.1 Practice Round	37
3.4.2 Round 1	38
3.4.3 Round 2	39
3.4.4 Round 3	39
3.4.5 Round 4	40
3.5 Evaluation of Educator Panel Meeting	40
3.6 Impact Data, Articulation, and Benchmarking	40
3.6.1 Impact Data	40
3.6.2 Articulation	44
3.6.3 Benchmark Data	47
4. REACTOR PANEL MEETING	50
5. PUBLIC INPUT WORKSHOPS	55
6. FINALIZATION OF ACHIEVEMENT STANDARDS	56
7. THE STANDARD SETTING RESULTS	57
8. REFERENCES	60

LIST OF APPENDICES

- Appendix A: Achievement Level Descriptions
- Appendix B: “Just Barely” Descriptions
- Appendix C: Agendas
- Appendix D: Security Plan
- Appendix E: Non-Disclosure Agreement
- Appendix F: Standard Setting Materials
- Appendix G: Educator Panel Demographic Information Sheets
- Appendix H: Reactor Panel Demographic Information Sheets
- Appendix I: Educator Panel Readiness Forms
- Appendix J: Reactor Panel Cut Score Recording Forms
- Appendix K: Educator Panel Evaluation Forms
- Appendix L: Reactor Panel Evaluation Forms
- Appendix M: Achievement Level Descriptions for NAEP, TIMSS, and PISA
- Appendix N: Past FCAT 2.0 data from 2011 to 2014
- Appendix O: Results of Round 1
- Appendix P: Results of Round 2
- Appendix Q: Results of Round 3
- Appendix R: Results of Round 4
- Appendix S: Results of Reactor Panel
- Appendix T: Impact Data Provided to the Educator Panel
- Appendix U: Impact Data Provided to the Reactor Panel
- Appendix V: Summary of Evaluations by the Educator Panel
- Appendix W: Summary of Evaluations by the Reactor Panel

LIST OF TABLES

Table 1: Achievement Levels and Achievement Standards	14
Table 2: Achievement Level Descriptions Panel.....	15
Table 3: Composition of Educator Panelists (Gender, Ethnicity, Occupations)	17
Table 4: Professional Experience of Educator Panelists	18
Table 5: Demographic Information of Educator Panelists (District Size, Community, and Region)	19
Table 6: Educator Panel in Various Occupations (by Committee).....	20
Table 7: Educator Panel Gender and Race/Ethnicity (by Committee)	21
Table 8: Educator Panel, Years Professional.....	22
Table 9: Educator Panel, Years Teaching Experience.....	23
Table 10: Educator Panel, Years Teaching Experience in Assigned Subject/Grade.....	24
Table 11: Educator Panel Type of Community	25
Table 12: Educator Panel Region of the State	26
Table 13: Educator Panel District Size.....	27
Table 14: Educator Panel Configuration	28
Table 15: Educator Panel Meeting, Attendees from FDOE and TDC.....	29
Table 16: Other AIR Staff Members in Attendance.....	30
Table 17: Summary of Required Meeting Space.....	32
Table 18: Sample Feedback from Round 1	38
Table 19: Theta to Scale Score Transformation Equations	41
Table 20: Impact Data (Grade 3 ELA)	42
Table 21: Percent in Florida at Each NAEP Standard.....	48
Table 22: Percent in Florida at each TIMSS International Benchmark.....	48
Table 23: Percent in Florida Reaching Each PISA Level.....	49
Table 24: Reactor Panel Attendees.....	50
Table 25: Reactor Panel Demographics.....	51
Table 26: Achievement Level Cuts in ELA from the Educator Panel.....	57
Table 27: Achievement Level Cuts in Mathematics from the Educator Panel.....	57
Table 28: Achievement Level Cuts in EOC Assessments from the Educator Panel.....	57
Table 29: Achievement Level Cuts in ELA from the Reactor Panel.....	58
Table 30: Achievement Level Cuts in Mathematics from the Reactor Panel.....	58
Table 31: Achievement Level Cuts in EOC Assessments from the Reactor Panel.....	58
Table 32: Achievement Level Cuts in ELA from the Commissioner.....	58
Table 33: Achievement Level Cuts in Mathematics from the Commissioner.....	59
Table 34: Achievement Level Cuts in EOC Assessments from the Commissioner.....	59

LIST OF FIGURES

Figure 1: Illustration of Standard Setting Process Used in Florida 2

Figure 2: ELA Achievement Level Cut Scores on the FSA Score Scale 4

Figure 3: ELA Percent Students At and Above Each Achievement Level..... 5

Figure 4: ELA Percent Students in Each Achievement Level..... 5

Figure 5: Mathematics Achievement Level Cut Scores on the FSA Score Scale 6

Figure 6: Mathematics Percent Students At and Above Each Achievement Level..... 6

Figure 7: Mathematics Percent Students in Each Achievement Level 7

Figure 8: End-of-Course (EOC) Achievement Level Cut Scores on the FSA Scale Score..... 7

Figure 9: End-of-Course (EOC) Percent Students At and Above Each Achievement Level..... 8

Figure 10: End-of-Course (EOC) Percent Students in Each Achievement Level 8

Figure 11: ELA Achievement Level Cut Scores on the FSA Score Scale 9

Figure 12: ELA Percent Students At and Above Each Achievement Level..... 9

Figure 13: ELA Percent Students in Each Achievement Level..... 10

Figure 14: Mathematics Achievement Level Cut Scores on the FSA Score Scale 10

Figure 15: Mathematics Percent Students At and Above Each Achievement Level..... 11

Figure 16: Mathematics Percent Students in Each Achievement Level..... 11

Figure 17: End-of-Course (EOC) Achievement Level Cut Scores on the FSA Score Scale..... 12

Figure 18: EOC Percent Students at and Above Each Achievement Level 12

Figure 19: EOC Percent Students in Each Achievement Level..... 13

Figure 20: “Just Barely” Descriptions 35

Figure 21: Bookmark Placement 36

Figure 22: Variability in Panelists’ Ratings..... 39

Figure 23: Articulation..... 44

Figure 24: Round 1 Feedback to Panelists..... 45

Figure 25: Articulation Feedback Given to Panelists 47

Figure 26: Impact Data for Reactor Panel: Percent Students in each Achievement Level
for ELA..... 53

Figure 27: Impact Data for Reactor Panel: Percent Students in each Achievement Level for
Mathematics..... 53

Figure 28: Impact Data for Reactor Panel: Percent Students in each Achievement Level
for EOC..... 54

1. EXECUTIVE SUMMARY

The Florida Standards Assessments (FSA) were designed to measure students' mastery of Florida's educational standards in English Language Arts (ELA), Mathematics, and End-of-Course (EOC) subjects (Algebra 1, Algebra 2, and Geometry). The FSA was primarily delivered as an online, fixed-form assessment; however, paper forms were universally administered to students in grades 3 and 4, and paper accommodated versions of grades 5–10 ELA and grades 5–8 Mathematics, as well as EOC assessments, were available to students who required them according to their Individual Education Plans (IEPs) or 504 Plans. The first administration of the FSA occurred in spring 2015, fulfilling the requirement, per Section 1008.22, Florida Statutes, to assess the academic achievement of Florida's public school students.

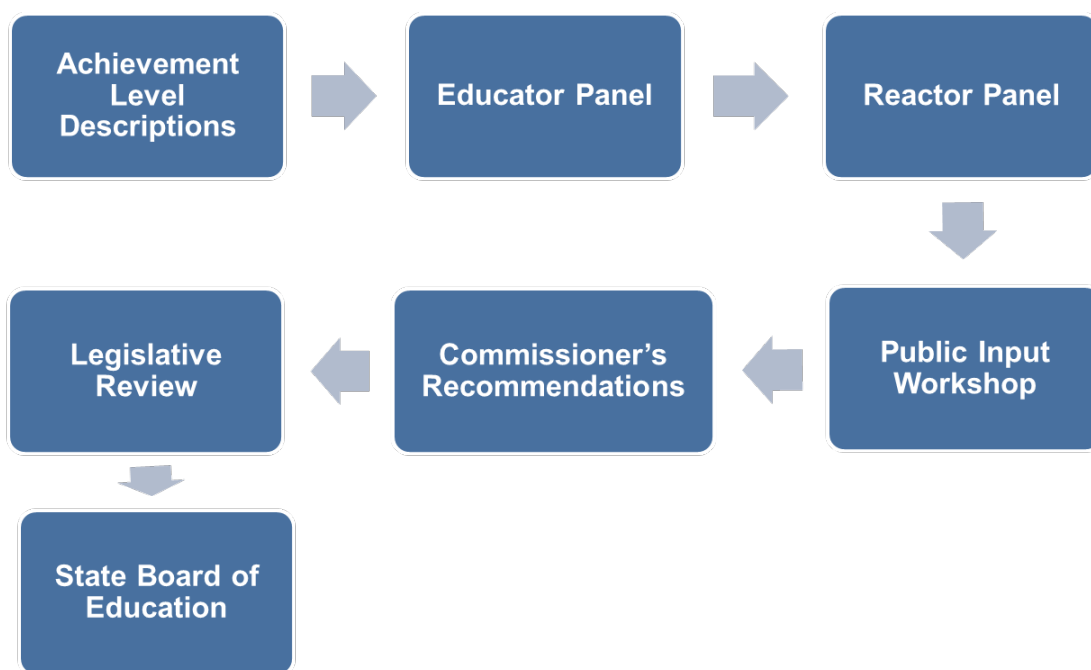
The FSA is aligned to the Florida Standards, which were approved by the Florida State Board of Education on February 18, 2014, to be the educational standards for all public schools in the state. The Florida Standards were implemented to establish rigorous expectations, with the goal of challenging and motivating Florida's students to acquire stronger critical thinking, problem solving, and communications skills. With the establishment of new standards and a new assessment, it is necessary to define achievement levels to effectively quantify student achievement on the new assessment. Standard setting is an empirical means of identifying achievement level cut scores to delineate these established levels of mastery.

The purpose of this technical report is to document the process and the results from the standard setting meetings, which were held to establish the cut scores for the FSA ELA, Mathematics, and EOC assessments. The Florida standard setting was a multi-stage process, as illustrated in Figure 1. The major sequence of events during standard setting was as follows:

1. It was first necessary to develop policy definitions for each of the achievement levels in all grades and subjects. These policy definitions provided overall guidance on the policy goals of each achievement level 1-5 and served as a basis for all stages of the process.
2. Achievement Level Descriptions (ALDs) were then developed to describe expectations of the student at each achievement level.
3. The Educator Panel meeting was held to recommend cut scores for each achievement level, with a diverse assembly of approximately 16–21 panelists assigned to each given grade/subject, with a total of 302 participants. The panel comprised a large body of experienced educators, most nominated by their school district superintendents, knowledgeable about educational standards, and experienced with the examinee population and sub-populations. The panel reviewed test content and made four rounds of cut score recommendations. AIR's web-based standard setting tool was used to collect the actual bookmarks (see Section 3.3.3 for detailed explanation of the Bookmark method) recorded by the panelists.
4. Following the Educator Panel, a Reactor Panel of community, business, and policy leaders was convened to review the established cut scores and to then potentially make additional recommendations based on a more extensive breakdown of impact data.

5. Public input on the panel recommendations was collected via workshops that were conducted statewide, open to all, and also made available online, as was a feedback survey.
6. Next, the Commissioner of Education considered the collection of feedback from the Educator Panel, the Reactor Panel, and the public workshops to make her recommendations, which were provided to the legislature and made publicly available on the FDOE website.
7. Finally, the Commissioner’s recommendations, as well as the information from the Educator Panel, the Reactor Panel, the public workshops, and legislative review were shared with the State Board of Education for consideration in their final decision of the adoption of the cut scores. On January 6, 2016, during the State Board of Education meeting, the State Board voted in favor of the Commissioner’s recommendations.

Figure 1: Illustration of Standard Setting Process Used in Florida



Below are definitions of important terms used throughout this document.

Standard Setting—A process of determining cut scores on a test that are associated with levels of achievement.

Achievement Level—Ranges across the test scale that represent degrees of mastery of the Florida Standards. There are five ranges: Level 1 (inadequate), Level 2 (below satisfactory), Level 3 (satisfactory), Level 4 (above satisfactory), and Level 5 (mastery).

Achievement Standard—A cut score on the test that indicates the minimum scaled score required to reach an achievement level. The four scaled score cut scores for each subject/grade will be associated with Level 2 (below satisfactory), Level 3 (satisfactory), Level 4 (above satisfactory), and Level 5 (mastery).

Cut Score—An FSA scaled score on the test that separates a lower achievement level from the next higher achievement level.

Policy Achievement Level Description (Policy ALD)—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.

Range Achievement Level Description (Range ALD)—A description of what students should know and be able to do throughout the range of an achievement level. In the FSA standard setting, ALDs were provided for Level 2, Level 3, Level 4, and Level 5. It was decided not to provide an ALD for Level 1 because there is no cut score established for Level 1.

Detailed Range ALD—ALD descriptions are provided for every content standard.

Summary Range ALD—ALDs are summarized across content standards.

“Just Barely” Achievement Level Description (Just Barely ALD)—A description of the lowest level of knowledge and skills required of a student to be considered in an achievement level. Just Barely ALDs are also called Target ALDs.

Reporting Achievement Level Description (Reporting ALD)—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 Summary Range ALDs.

Bookmark Method—A method of standard setting where a panel of educators marks the page in an ordered item booklet that best represents an achievement standard.

Response Probability—The requisite conditional probability that the just barely proficient student will correctly answer an item in the ordered item booklet (usually set to a probability of .67).

Ordered Item Booklet—A booklet of items proportional to the test blueprint where the items have been ordered by difficulty from easy to hard.

Educator Panel—A group of trained and experienced educators who recommend cut scores that best represent achievement standards/levels.

Reactor Panel—A group of stakeholders with diverse viewpoints who review the recommendations of the Educator Panel and suggest revisions based on policy considerations.

Vertical Articulation—Cut scores that are incrementally higher in higher grades with no unexpected dips or spikes.

Impact Data—Statewide data that show what percentage of students are impacted by various cut scores.

Benchmark Data—Data that show how the internal state achievement standards compare with important external standards.

The subsequent sections of this report describe the process for recommending achievement standards. This extensive, collaborative process is intended to result in cut scores that are

- content-referenced, because they were based on a rigorous application of the Florida Standards;
- articulated across grades, with the establishment of a vertical scale based on student performance data;
- reasonable, as they were based on informed judgments of the Educator and Reactor Panel experts;
- credible, because a diverse group of panelists followed a rigorous and well-supported standard setting procedure; and
- benchmarked against empirical external indicators.

Recommended Achievement Standards

The achievement level cut scores recommended by the Educator Panel and the Reactor Panel are provided below.

Results from Educator Panel

Figure 2 provides the achievement level cut scores for ELA on the FSA score scale. Figure 3 provides the percentage of students at and above each achievement level cut in ELA, and Figure 4 presents the percentage of students in each achievement level for ELA. The percentage of students within the state who meet or exceed each potential achievement level was estimated using spring 2015 results data (baseline operational FSA administration). In other words, the percentages depicted in Figure 2 and Figure 3 reflect how students in the spring 2015 administration would have performed if the Educator Panel’s recommendations had been in effect.

Figure 2: ELA Achievement Level Cut Scores on the FSA Score Scale

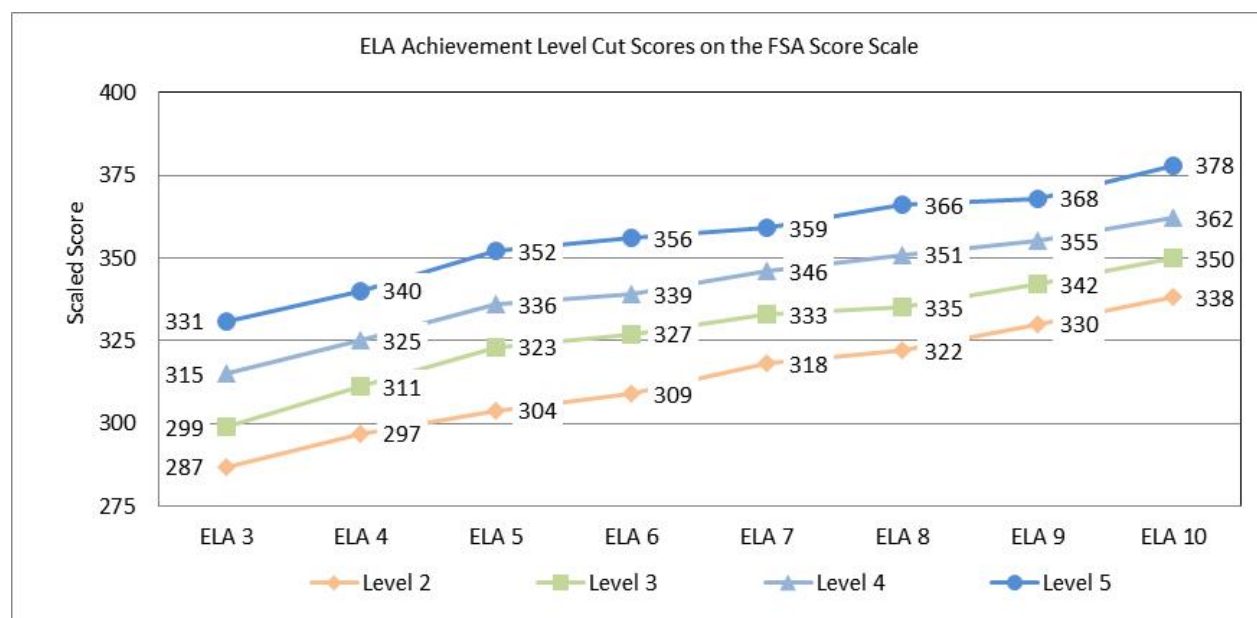


Figure 3: ELA Percent Students At and Above Each Achievement Level

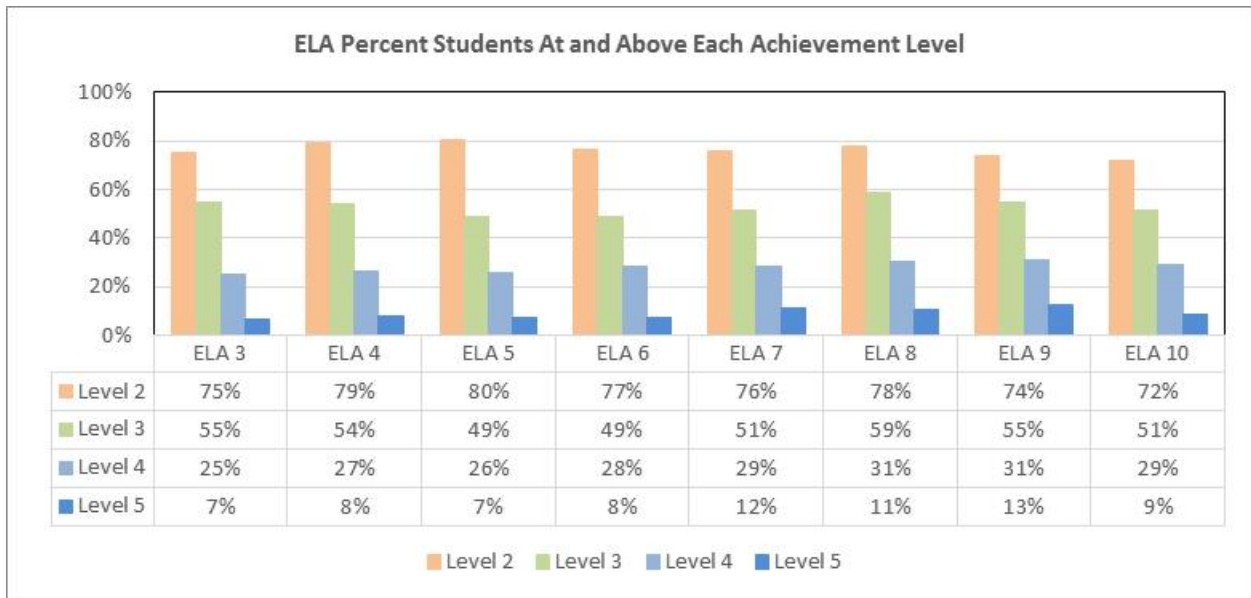


Figure 4: ELA Percent Students in Each Achievement Level

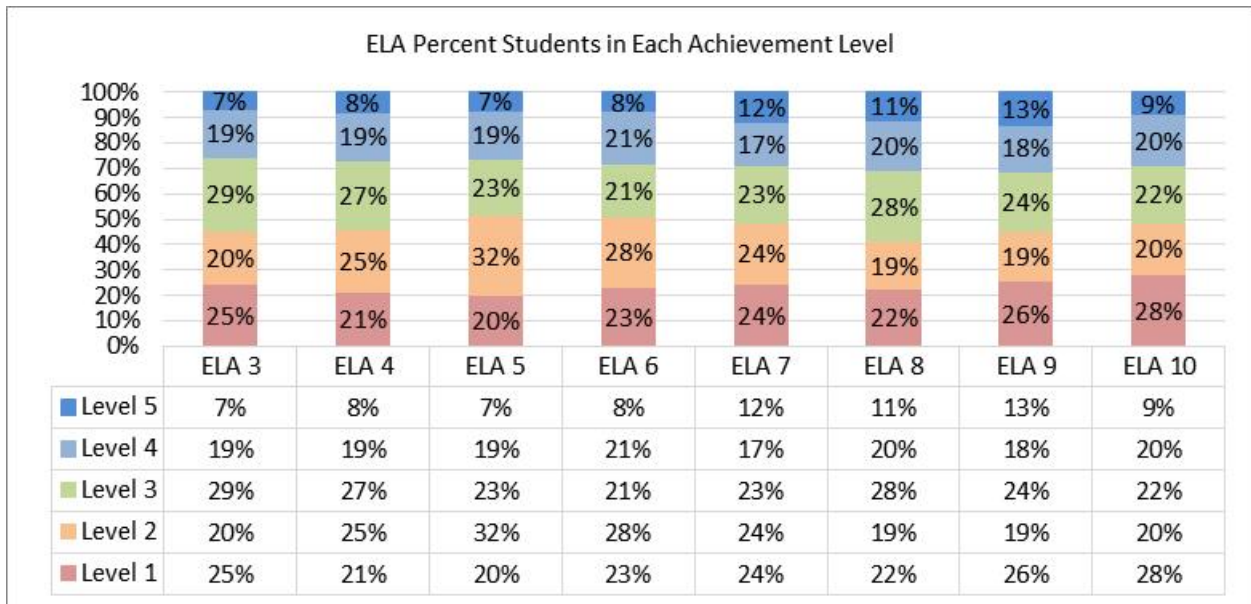


Figure 5 provides the achievement level cut scores for Mathematics on the FSA score scale. Figure 6 displays the percentage of students at and above each achievement level in Mathematics, and Figure 7 presents the percentage of students in each achievement level for Mathematics. These percentages are based on spring 2015 FSA results data.

Figure 5: Mathematics Achievement Level Cut Scores on the FSA Score Scale

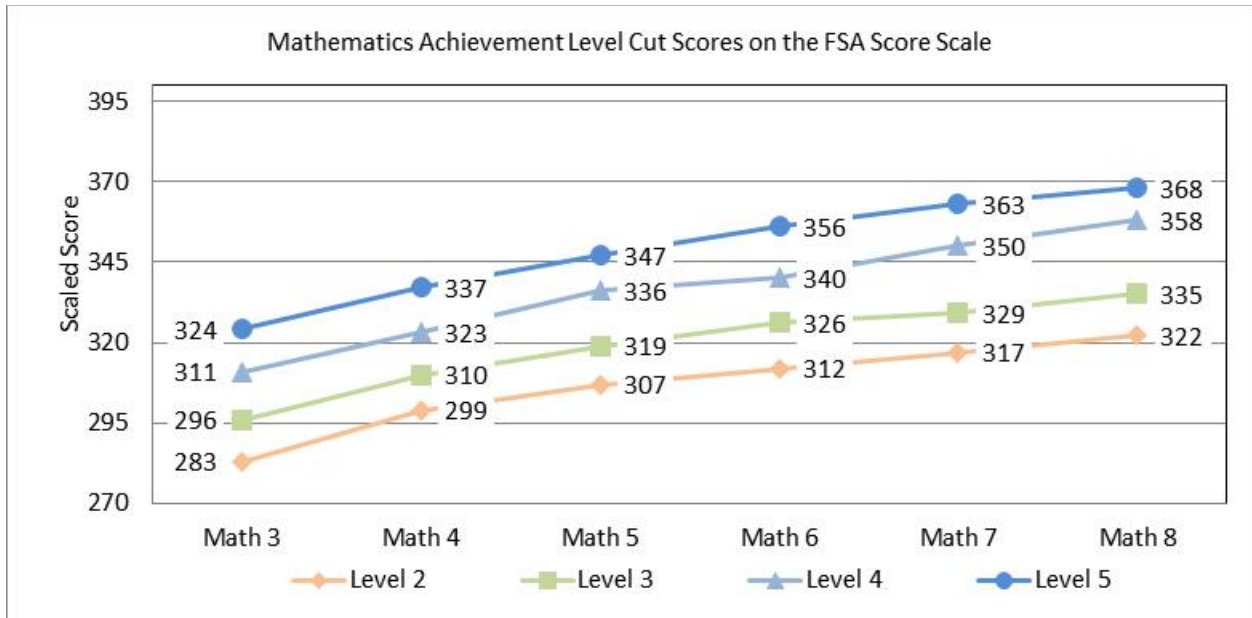


Figure 6: Mathematics Percent Students At and Above Each Achievement Level

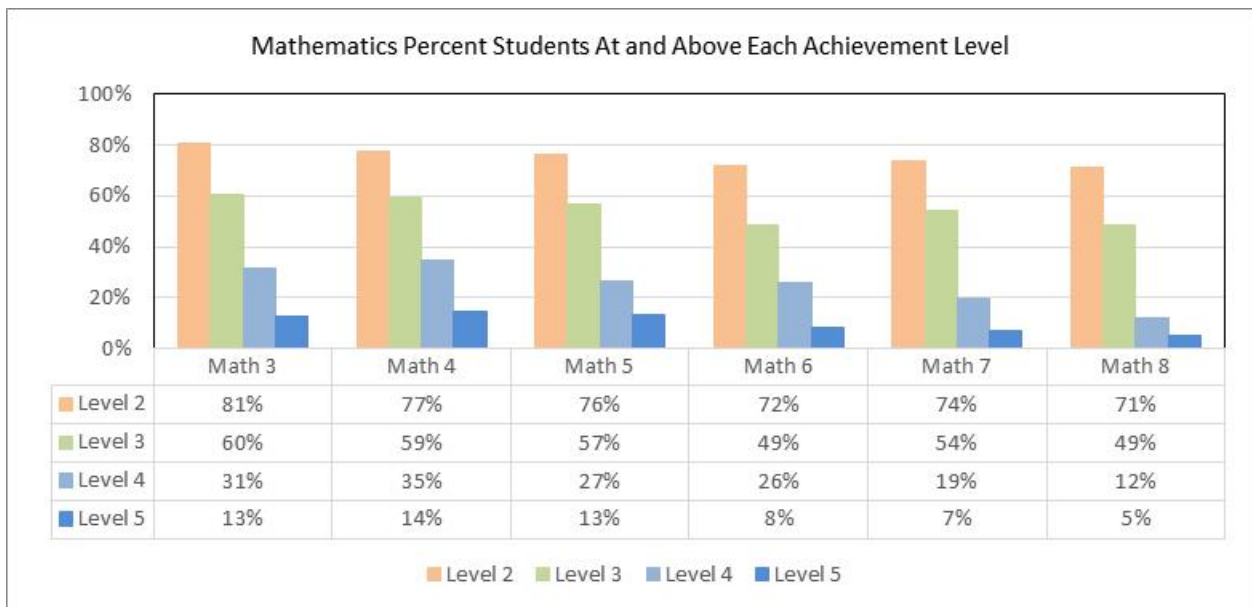


Figure 7: Mathematics Percent Students in Each Achievement Level

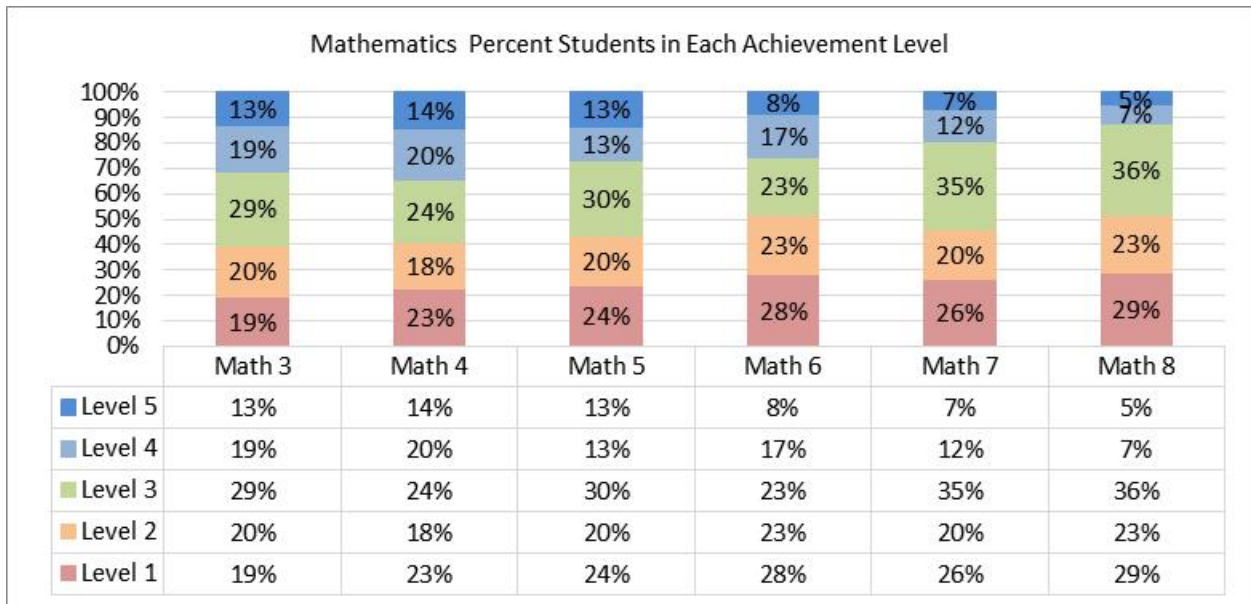


Figure 8 displays the achievement level cut scores for EOC assessments on the FSA score scale.

Figure 9 provides the percentage of students at and above each achievement level, and Figure 10 presents the percentage of EOC students in each achievement level. These percentages are based on spring 2015 FSA results data.

Figure 8: End-of-Course (EOC) Achievement Level Cut Scores on the FSA Scale Score

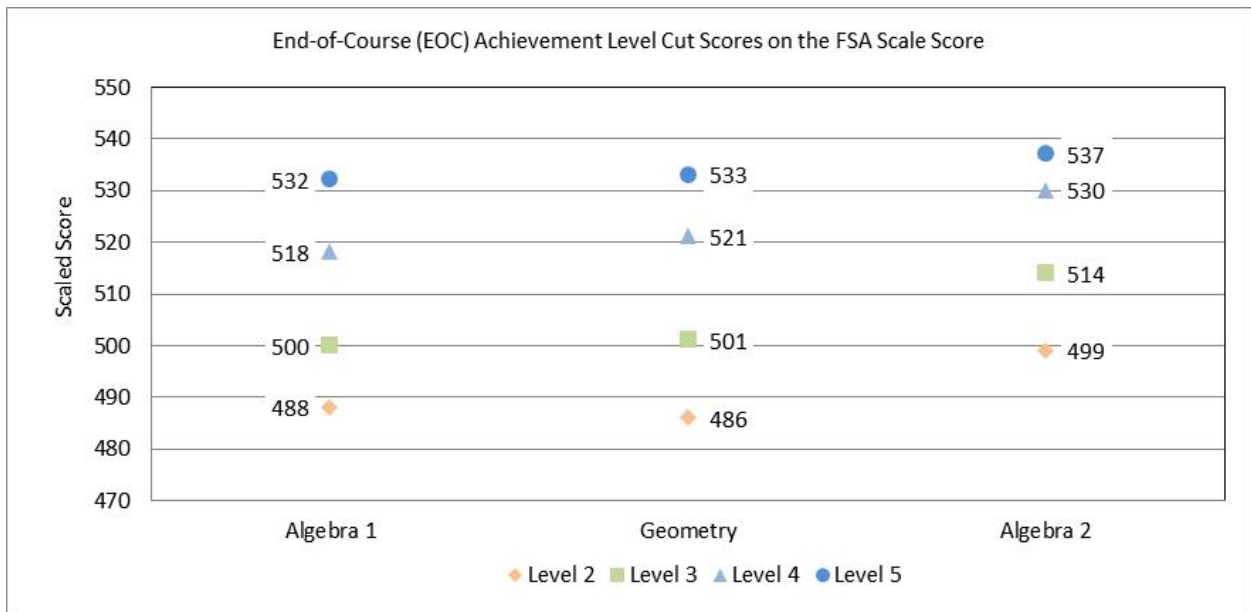


Figure 9: End-of-Course (EOC) Percent Students At and Above Each Achievement Level

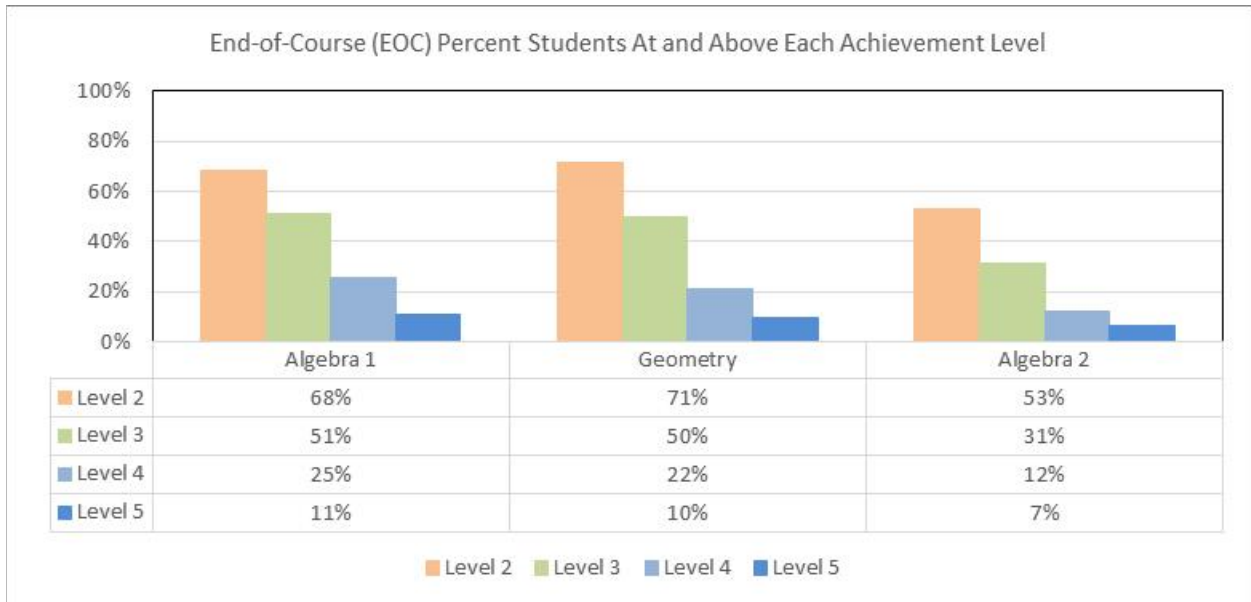
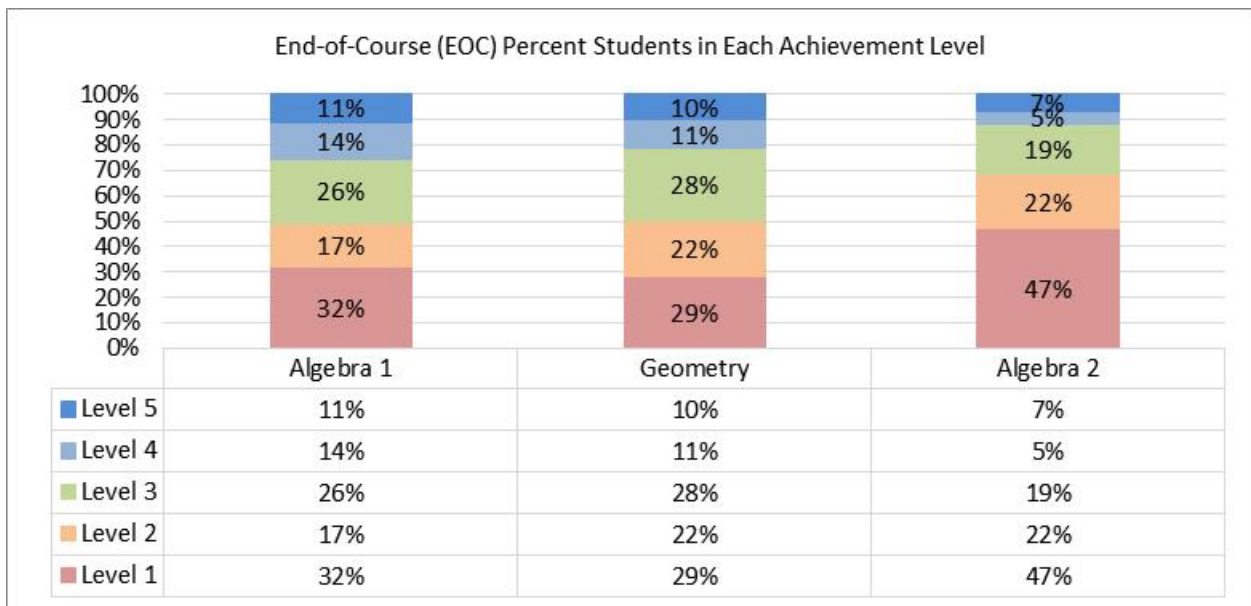


Figure 10: End-of-Course (EOC) Percent Students in Each Achievement Level



Results from Reactor Panel

This section provides the cut scores for each grade and subject recommended by the Reactor Panel. Figure 11 provides the achievement level cut scores for ELA on the FSA score scale. Figure 12 provides the percentage of students at and above each achievement level in ELA. Figure 13 presents the percentage of students in each achievement level for ELA. Similarly, the

results for Mathematics are provided in Figure 14 through Figure 16, and those for EOCs are provided in Figure 17 through Figure 19.

Figure 11: ELA Achievement Level Cut Scores on the FSA Score Scale

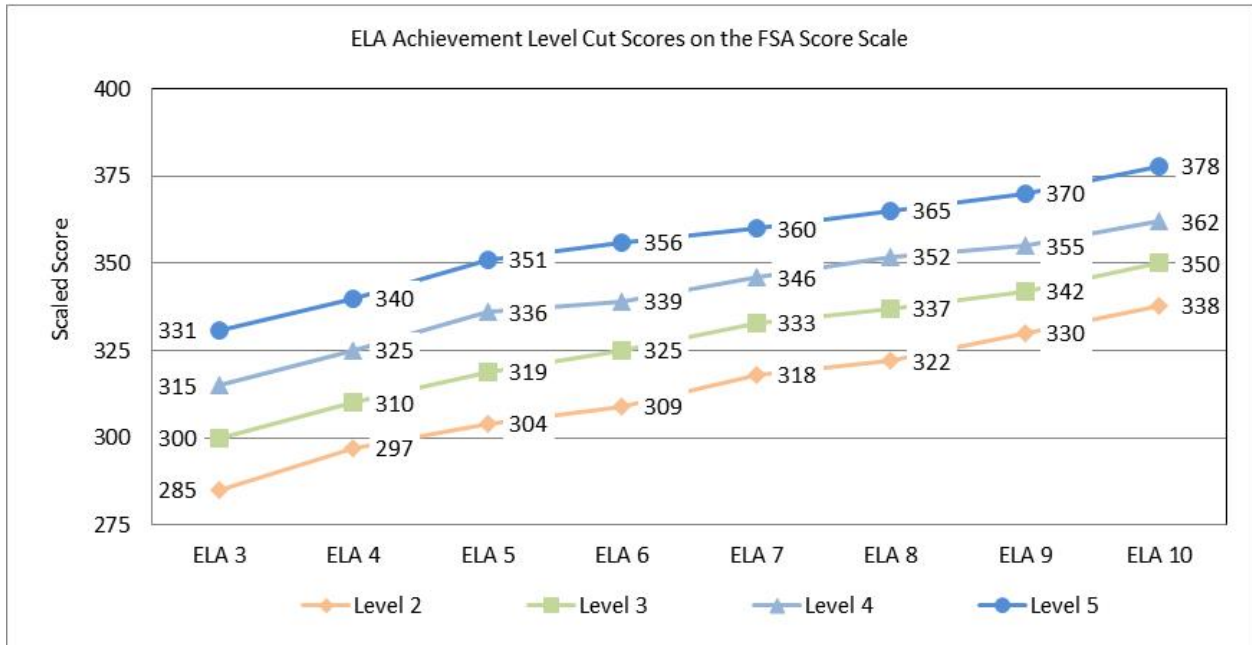


Figure 12: ELA Percent Students At and Above Each Achievement Level

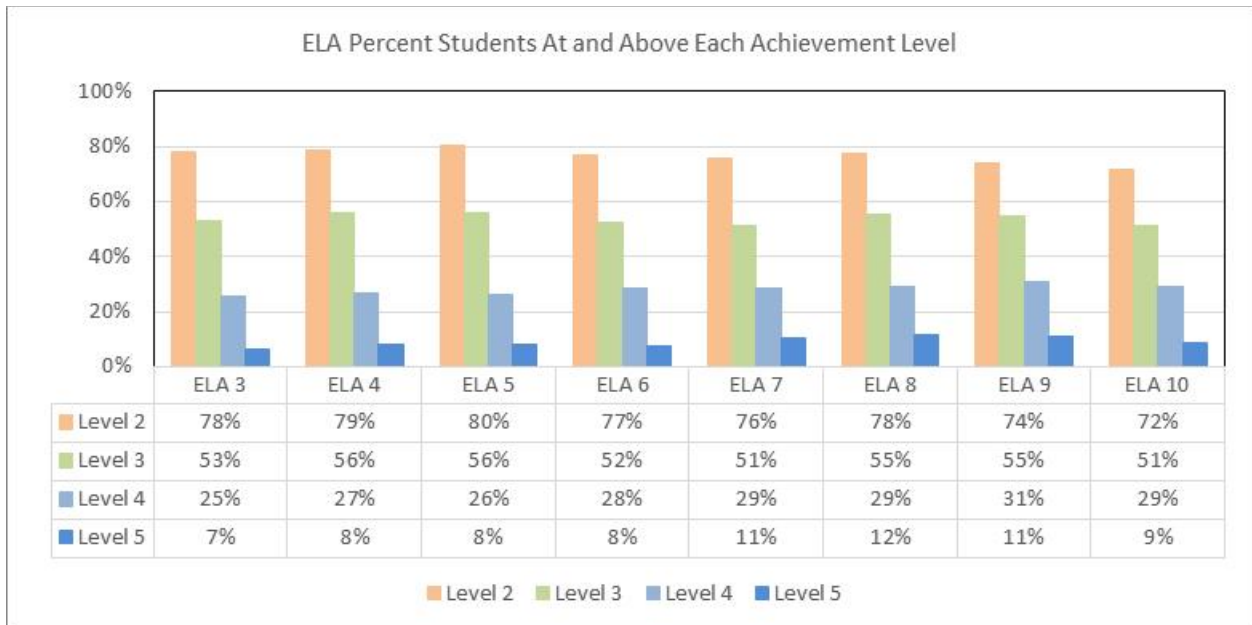


Figure 13: ELA Percent Students in Each Achievement Level

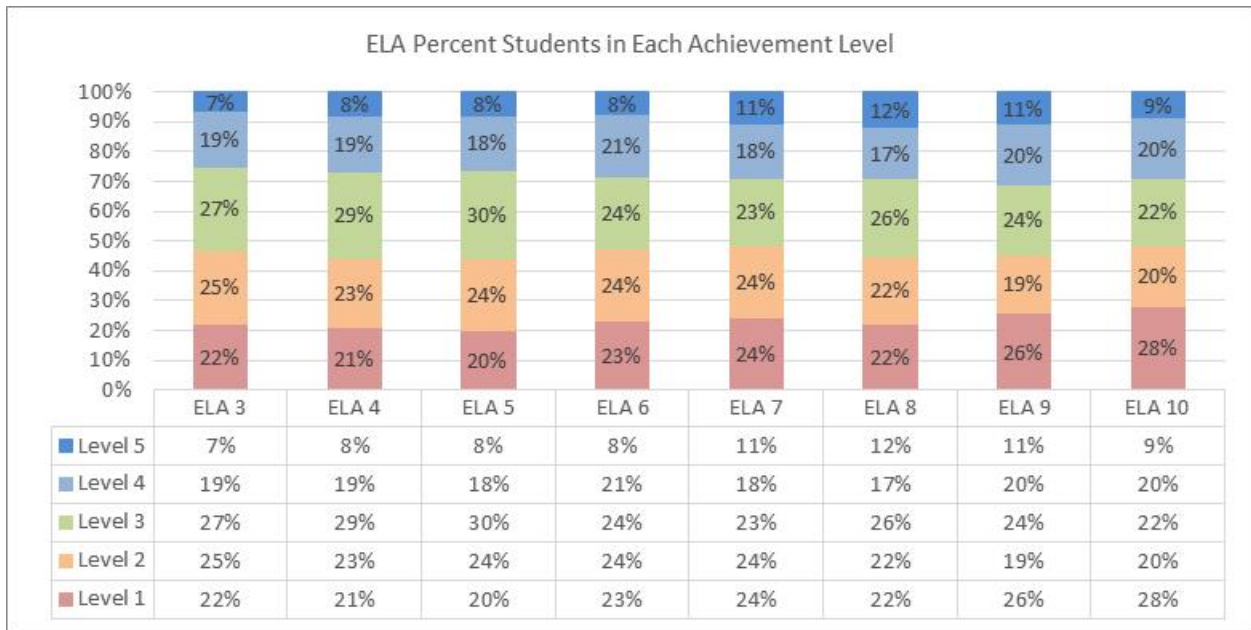


Figure 14: Mathematics Achievement Level Cut Scores on the FSA Score Scale

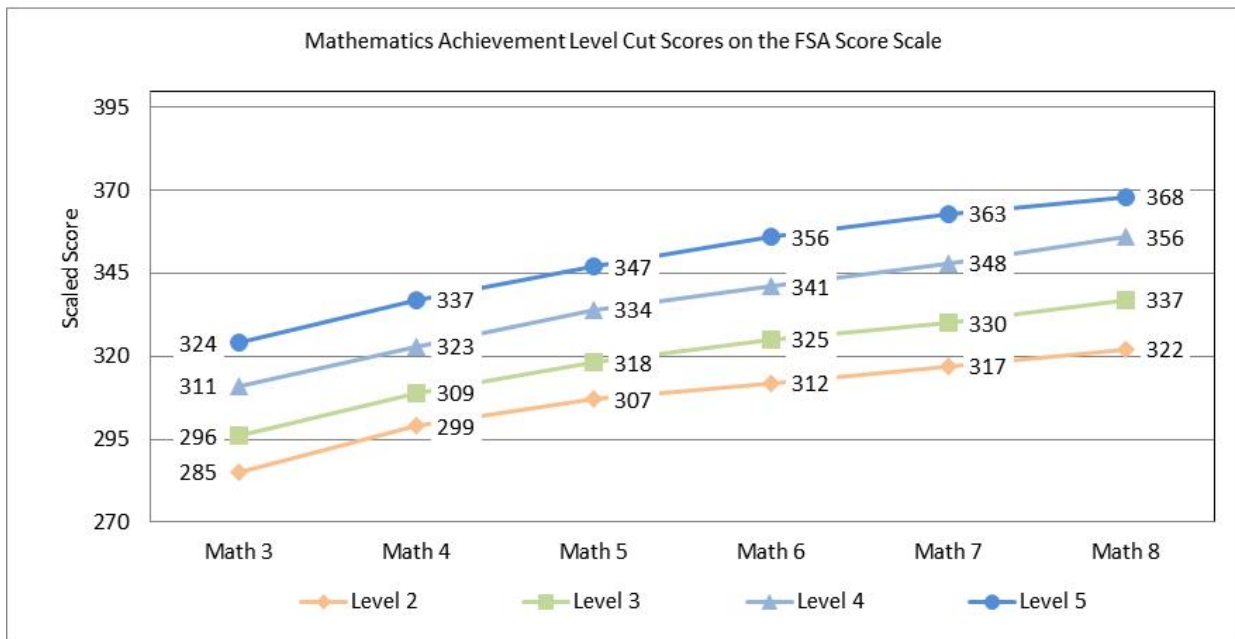


Figure 15: Mathematics Percent Students At and Above Each Achievement Level

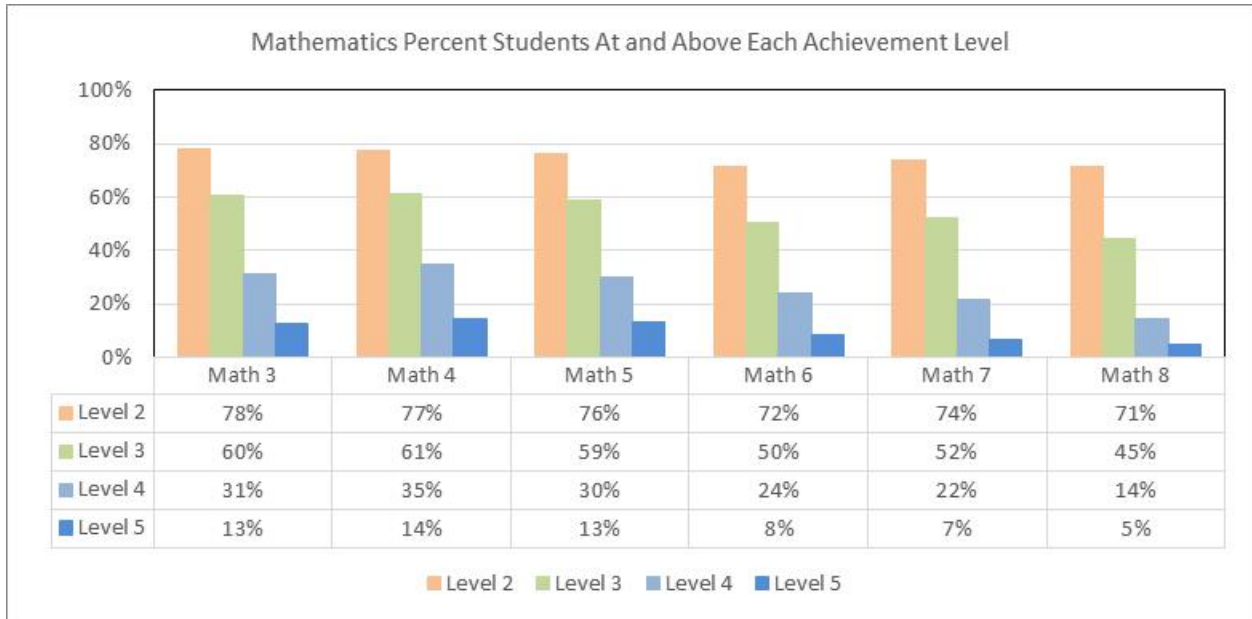


Figure 16: Mathematics Percent Students in Each Achievement Level

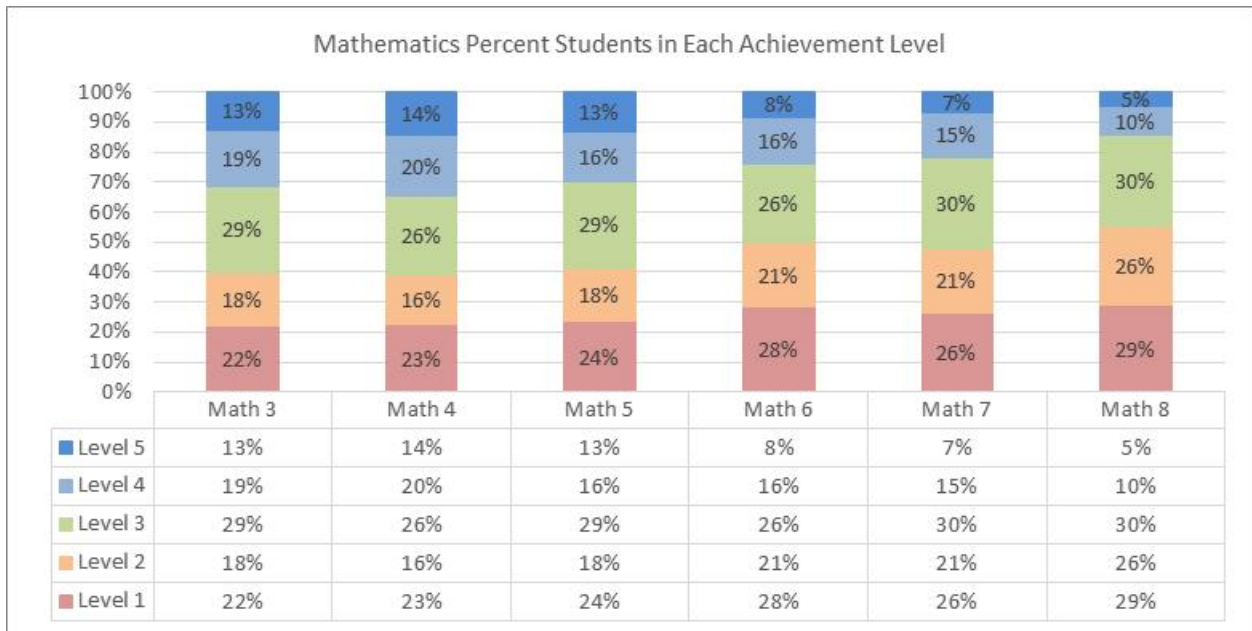


Figure 17: End-of-Course (EOC) Achievement Level Cut Scores on the FSA Score Scale

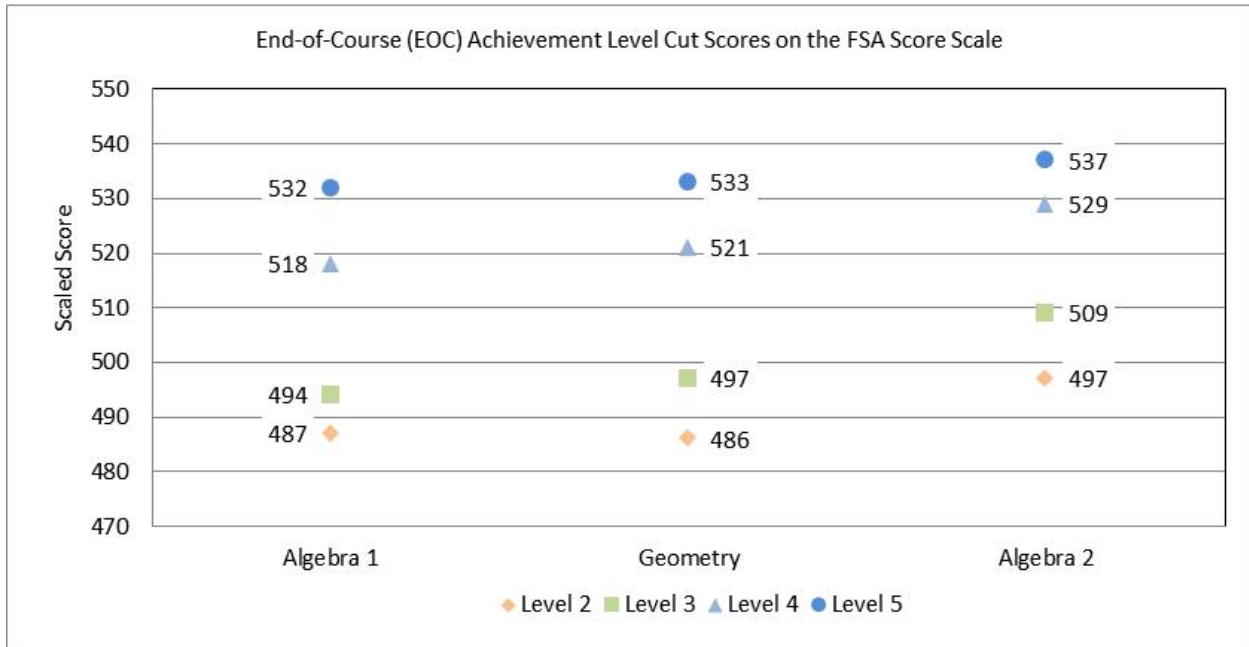


Figure 18: EOC Percent Students at and Above Each Achievement Level

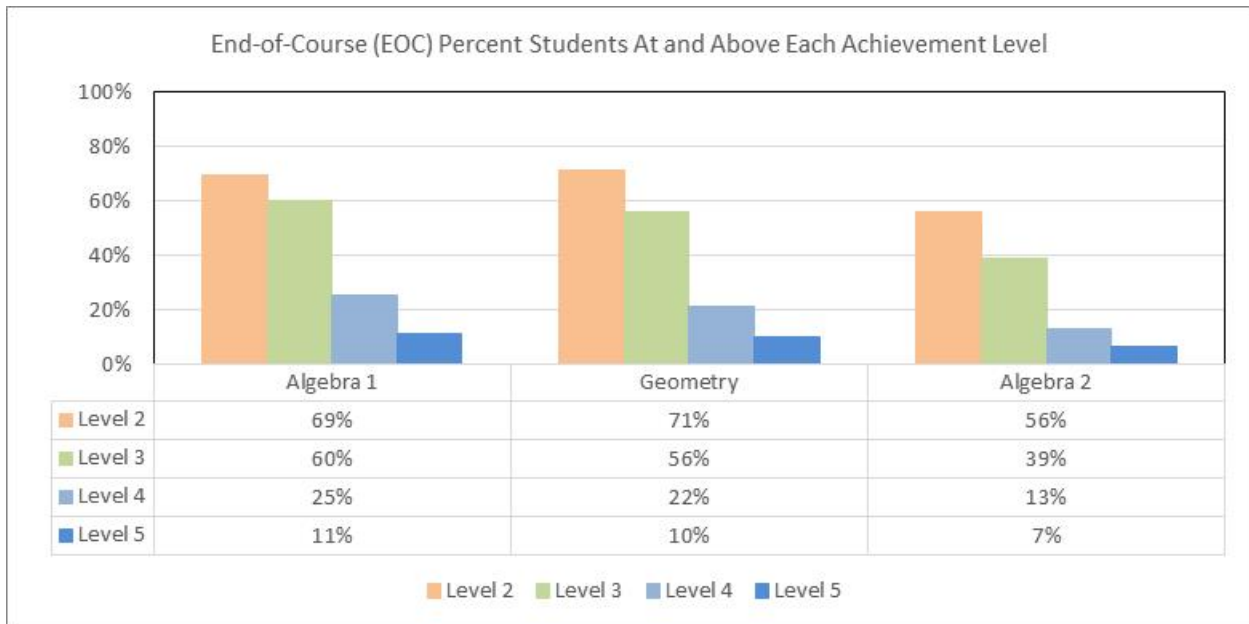
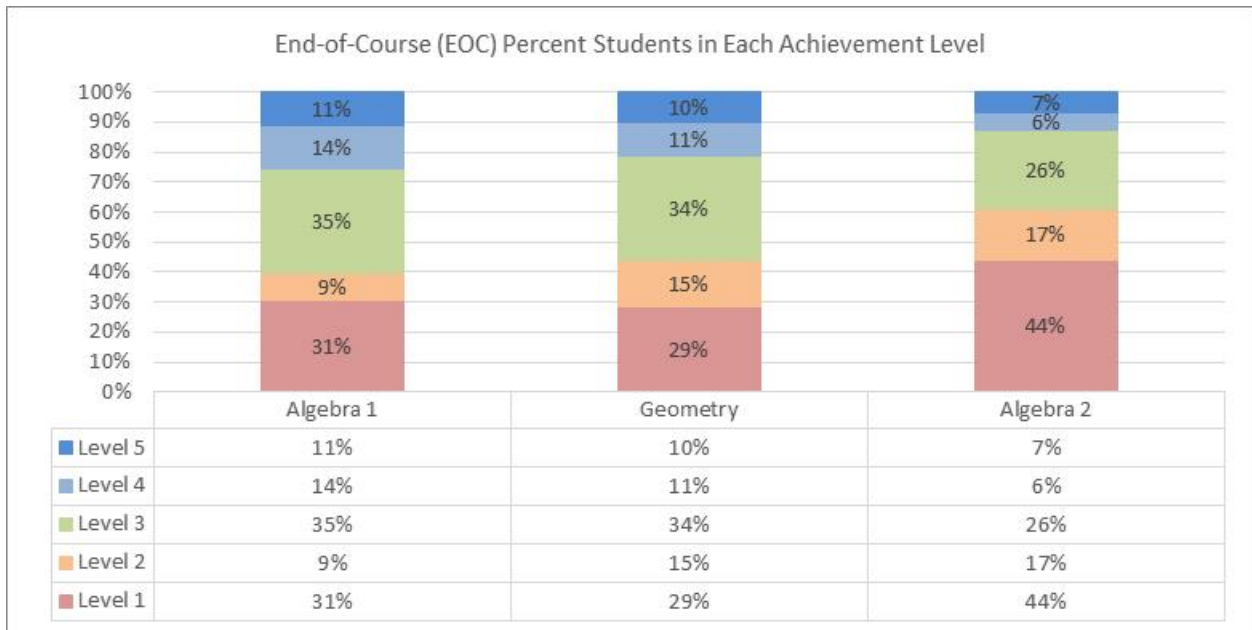


Figure 19: EOC Percent Students in Each Achievement Level



2. BACKGROUND

As detailed in Volume 2, the Florida Standards Assessments were aligned to the Florida Standards test blueprints, academic content standards, and reporting categories; these Florida Standards can also be accessed at <http://www.flstandards.org/>. In order to measure student achievement on the FSA, standard setting is necessary as the means of identifying cut scores on the FSA score scale to indicate the boundaries of the five student achievement levels.

Florida used the established Bookmark method of standard setting (Mitzel, Lewis, Patz & Green, 2001), which is the most common procedure used around the country. In this process, panelists of Florida educators reviewed test items in tandem with the corresponding Florida Content Standards and Achievement Level Descriptions (ALDs), and they then recommended cut scores, which are also often referred to as achievement standards. While some standard setting methods employ normative techniques, the Bookmark method utilizes empirically tested techniques that emphasize standard criteria and the expected skills of students. This Bookmark standard setting process was described in a Standard Setting Plan submitted to FDOE and reviewed and approved by FDOE prior to its implementation. A preliminary version of the plan was also presented to the Florida Technical Advisory Committee (TAC) prior to the Educator Panel and Reactor Panel standard setting meetings.

The content standards are the primary consideration when using the Bookmark method, where operational test items for a given grade or subject are ordered by difficulty, and panelists make judgments about which items students performing at each achievement level would be able to complete correctly. This method is further outlined in Section 3.

2.1 ACHIEVEMENT LEVELS

The cut scores, or achievement standards, established in the standard setting process represent the lowest boundary of each achievement level, which are defined in the Achievement Level Policy Definitions, as follows:

Table 1: Achievement Levels and Achievement Standards

Level 1	Level 2	Level 3	Level 4	Level 5
Students at this level demonstrate an inadequate level of success with the challenging content of the <i>Florida Standards</i> .	Students at this level demonstrate a below satisfactory level of success with the challenging content of the <i>Florida Standards</i> .	Students at this level demonstrate a satisfactory level of success with the challenging content of the <i>Florida Standards</i> .	Students at this level demonstrate an above satisfactory level of success with the challenging content of the <i>Florida Standards</i> .	Students at this level demonstrate mastery of the most challenging content of the <i>Florida Standards</i> .

Achievement Level Descriptions (ALDs) define the intellectual constructs, knowledge, and skills based on the Florida Standards that are expected of a student at each of these levels. Once the

standard setting process is finalized, the cut scores and score ranges for each achievement level for each test grade and subject will be established.

These levels are used for reporting to parents, teachers, and schools, as well as for federal reporting. Moreover, because student progress from grade to grade is a central interest for any given assessment, these cut scores and the levels of achievement they represent must increase incrementally from grade to grade. That is, at a reasonable rate of progress, it should not be expected that a student who reached Level 3 in one year would be labeled a Level 2 student the following year. It would be difficult to interpret results in which large numbers of students demonstrate irregular, dramatic changes in achievement levels, when their progress is realistically consistent with teacher and program expectations.

2.2 ACHIEVEMENT LEVEL DESCRIPTIONS

The Achievement Level Descriptions (ALDs) specify the learning expectations of the achievement level content standards for all grades and are based on the Florida Standards. Range ALDs define the knowledge, skills, and processes that examinees at a particular achievement level are expected to possess for each content area. During the entire standard setting process, these content-aligned descriptions were used to inform judgments of the placement of the cut scores.

The Range ALDs were drafted by the Center for Assessment and reviewed and revised by the ALD panel of Florida educators in a meeting held April 28–May 1, 2015, at the Hotel Duval in Tallahassee, Florida. Summaries of the Range ALDs are provided in Appendix A. Each breakout room of the ALD panel was facilitated by a representative from the Department’s Test Development Center (TDC) and an AIR content specialist. Additional AIR staff members were in attendance, including Gary Phillips (VP), Kevin Dwyer (ELA), Meg McMahon (Mathematics), Chris Paskoff (Mathematics), and Diana Reed (ELA and Florida Item Development Manager). The number of participants for the ALD panel is displayed in Table 2.

Table 2: Achievement Level Descriptions Panel

Participants per Breakout Room				
Breakout Room	Participants	TDC and AIR	TDC Staff	AIR Staff
ELA Grades 3–5	6	2	Elizabeth Tricquet	Allison Stingley
ELA Grades 6–7	8	2	Michelle Peddie	Brett Craycraft
ELA Grades 8 and 10	7	4	Gretchen Sims and Sally Rhodes	Natalie Rebentisch and Kelly Quinney
ELA Overall		2	Renn Edenfield	Meghan Mulhern
Mathematics Grades 3–5	6	2	Travis Barton	Alysa Kartee
Mathematics Grades 6–8	6	2	June Campbell	Jen Rubel
Mathematics EOCs	9	2	Terri Sebring	Kathy Sagris
Mathematics Overall		2	Chris Harvey	Maureen Font
Total	42	18		

3. EDUCATOR PANEL

During the week of August 31–September 4, 2015, the Florida Department of Education (FDOE) convened a diverse panel of more than 300 educators in Orlando, Florida. The purpose of this Educator Panel meeting was to recommend the cut scores for each achievement level of the Florida Standards Assessments. Achievement level cuts were recommended for English Language Arts (ELA) tests in grades 3–10, Mathematics tests in grades 3–8, and End-of-Course (EOC) assessments in Algebra 1, Algebra 2, and Geometry. The Educator Panel was comprised a large number of educators with expertise in the content areas of their respective committees as well as experience and understanding of students’ knowledge of those grade/subject-level standards. Using the Bookmark method, the Educator Panel recommended cut scores in four rounds, making judgments based on what students should know and be able to do at each achievement level in each grade and subject. The cut scores recommended at this meeting served as a foundation for all subsequent steps in the standard setting process. In the sections that follow are details regarding the attendees of the Educator Panel meeting, logistical preparation, trainings, articulation, impact data, and benchmarking.

3.1 EDUCATOR PANEL COMPOSITION

FDOE solicited superintendent nominations and selected the committee members for the Educator Panel. The Educator Panel consisted of 302 panelists recruited by FDOE from across the state. The recruiting plan for obtaining panelists for the standard setting meetings was designed to establish representative groups of panelists who would render informed, content-based recommendations to the state on the placement of the cut scores for each achievement level. Diverse groups of panelists for each individual grade and subject brought a wide range of perspectives and experience to the standard setting procedure, ensuring that the recommendations were thoughtful and representative of broad educational constituencies.

The demographic information of the panelists was collected using the forms provided in Appendix G, and the results from the demographic sheet are summarized in Table 3 through Table 5.

Table 3 provides the composition of the Educator Panel in terms of gender, race/ethnicity, and occupation. Table 4 summarizes the educator panelists’ years of experience in terms of both teaching experience and other professional experience, such as working as an administrator or specialist. Table 5 displays the summary of demographic information by district size, community, and region of educator panelists. Table 6 through Table 13 provide the disaggregated summary.

Table 3: Composition of Educator Panelists (Gender, Ethnicity, Occupations)

Gender	Total		ELA		Mathematics		EOC	
	N	Percentage	N	Percentage	N	Percentage	N	Percentage
Male	39	13%	10	7%	17	15%	12	24%
Female	260	86%	126	91%	95	85%	39	76%
NR*	3	1%	3	2%	0	0%	0	0%
<i>Total</i>	<i>302</i>	<i>100%</i>	<i>139</i>	<i>100%</i>	<i>112</i>	<i>100%</i>	<i>51</i>	<i>100%</i>
Race/Ethnicity	N	Percentage	N	Percentage	N	Percentage	N	Percentage
White	216	72%	103	74%	79	71%	34	67%
African American	40	13%	17	12%	17	15%	6	12%
Hispanic	25	8%	10	7%	9	8%	6	12%
Other	18	6%	6	4%	7	6%	5	10%
NR*	3	1%	3	2%	0	0%	0	0%
<i>Total</i>	<i>302</i>	<i>100%</i>	<i>139</i>	<i>100%</i>	<i>112</i>	<i>100%</i>	<i>51</i>	<i>100%</i>
Occupations	N	Percentage	N	Percentage	N	Percentage	N	Percentage
Teacher	89	29%	38	27%	31	28%	20	39%
Coach	52	17%	22	16%	25	22%	5	10%
Specialist	75	25%	31	22%	28	25%	16	31%
Administrator	33	11%	19	14%	11	10%	3	6%
Other	36	12%	22	16%	8	7%	6	12%
Two or more occupations	16	5%	6	4%	9	8%	1	2%
NR*	1	0%	1	1%	0	0%	0	0%
<i>Total</i>	<i>302</i>	<i>100%</i>	<i>139</i>	<i>100%</i>	<i>112</i>	<i>100%</i>	<i>51</i>	<i>100%</i>

*NR indicates *not reported*.

Table 4: Professional Experience of Educator Panelists

Years Professional	Total		ELA		Mathematics		EOC	
	N	%	N	%	N	%	N	%
Less than or equal to 1 year	66	22%	22	16%	29	26%	15	29%
1–5 years	114	38%	57	41%	39	35%	18	35%
6–10 years	52	17%	20	14%	23	21%	9	18%
11–15 years	32	11%	17	12%	10	9%	5	10%
16–20 years	12	4%	10	7%	2	2%	0	0%
21 or more years	17	6%	8	6%	6	5%	3	6%
NR	9	3%	5	4%	3	3%	1	2%
<i>Total</i>	<i>302</i>	<i>100%</i>	<i>139</i>	<i>100%</i>	<i>112</i>	<i>100%</i>	<i>51</i>	<i>100%</i>
Years Teaching	N	%	N	%	N	%	N	%
Less than or equal to 1 year	0	0%	0	0%	0	0%	0	0%
1–5 years	14	5%	8	6%	6	5%	0	0%
6–10 years	89	29%	40	29%	36	32%	13	25%
11–15 years	87	29%	43	31%	34	30%	10	20%
16–20 years	48	16%	24	17%	13	12%	11	22%
21 or more years	64	21%	24	17%	23	21%	17	33%
NR	0	0%	0	0%	0	0%	0	0%
<i>Total</i>	<i>302</i>	<i>100%</i>	<i>139</i>	<i>100%</i>	<i>112</i>	<i>100%</i>	<i>51</i>	<i>100%</i>
Years Teaching in Assigned Subject	N	%	N	%	N	%	N	%
Less than or equal to 1 year	21	7%	10	7%	10	9%	1	2%
1–5 years	100	33%	55	40%	41	37%	4	8%
6–10 years	92	30%	32	23%	36	32%	24	47%
11–15 years	43	14%	19	14%	17	15%	7	14%
16–20 years	29	10%	12	9%	5	4%	12	24%
21 or more years	10	3%	5	4%	2	2%	3	6%
NR	7	2%	6	4%	1	1%	0	0%
<i>Total</i>	<i>302</i>	<i>100%</i>	<i>139</i>	<i>100%</i>	<i>112</i>	<i>100%</i>	<i>51</i>	<i>100%</i>

Table 5: Demographic Information of Educator Panelists (District Size, Community, and Region)

District Size	Total		ELA		Mathematics		EOC	
	N	Percentage	N	Percentage	N	Percentage	N	Percentage
Large	132	44%	58	42%	49	44%	25	49%
Medium	80	26%	39	28%	31	28%	10	20%
Small	84	28%	38	27%	31	28%	15	29%
NR	6	2%	4	3%	1	1%	1	2%
<i>Total</i>	302	100%	139	100%	112	100%	51	100%
Community	N	Percentage	N	Percentage	N	Percentage	N	Percentage
Urban	96	32%	41	29%	40	36%	15	29%
Suburban	107	35%	52	37%	35	31%	20	39%
Rural	88	29%	39	28%	34	30%	15	29%
NR	11	4%	7	5%	3	3%	1	2%
<i>Total</i>	302	100%	139	100%	112	100%	51	100%
Region	N	Percentage	N	Percentage	N	Percentage	N	Percentage
Panhandle	50	17%	24	17%	18	16%	8	16%
Northeast	45	15%	22	16%	18	16%	5	10%
East Central	65	22%	31	22%	23	21%	11	22%
West Central	59	20%	29	21%	21	19%	9	18%
South	80	26%	32	23%	31	28%	17	33%
NR	3	1%	1	1%	1	1%	1	2%
<i>Total</i>	302	100%	139	100%	112	100%	51	100%

Table 6: Educator Panel in Various Occupations (by Committee)

Committee		Teacher	Coach	Specialist	Administrator	Other	Two or more occupations	NR
Gr. 3 Math	N	7	3	2	3	2	1	0
	%	39%	17%	11%	17%	11%	6%	0%
Gr. 4 Math	N	5	5	3	0	2	4	0
	%	26%	26%	16%	0%	11%	21%	0%
Gr. 5 Math	N	3	9	7	1	1	0	0
	%	14%	43%	33%	5%	5%	0%	0%
Gr. 6 Math	N	9	1	3	2	1	2	0
	%	50%	6%	17%	11%	6%	11%	0%
Gr. 7 Math	N	5	4	7	2	1	1	0
	%	25%	20%	35%	10%	5%	5%	0%
Gr. 8 Math	N	2	3	6	3	1	1	0
	%	13%	19%	38%	19%	6%	6%	0%
Algebra 1	N	6	3	7	1	2	0	0
	%	32%	16%	37%	5%	11%	0%	0%
Algebra 2	N	8	1	4	1	1	1	0
	%	50%	6%	25%	6%	6%	6%	0%
Geometry	N	6	1	5	1	3	0	0
	%	38%	6%	31%	6%	19%	0%	0%
Gr. 3 ELA	N	2	7	4	3	1	0	0
	%	12%	41%	24%	18%	6%	0%	0%
Gr. 4 ELA	N	5	1	3	3	4	3	0
	%	26%	5%	16%	16%	21%	16%	0%
Gr. 5 ELA	N	5	3	4	3	3	0	0
	%	28%	17%	22%	17%	17%	0%	0%
Gr. 6 ELA	N	4	4	5	4	4	0	0
	%	19%	19%	24%	19%	19%	0%	0%
Gr. 7 ELA	N	3	3	5	1	1	2	0
	%	20%	20%	33%	7%	7%	13%	0%
Gr. 8 ELA	N	6	3	1	0	4	1	1
	%	38%	19%	6%	0%	25%	6%	6%
Gr. 9 ELA	N	8	1	2	3	1	0	0
	%	53%	7%	13%	20%	7%	0%	0%
Gr. 10 ELA	N	5	0	7	2	4	0	0
	%	28%	0%	39%	11%	22%	0%	0%

Table 7: Educator Panel Gender and Race/Ethnicity (by Committee)

Committee		Gender			Race/Ethnicity				
		Male	Female	Gender NR	White	African American	Hispanic	Other	NR
Gr. 3 Math	N	0	18	0	12	2	3	1	0
	%	0%	100%	0%	67%	11%	17%	6%	0%
Gr. 4 Math	N	3	16	0	16	1	1	1	0
	%	16%	84%	0%	84%	5%	5%	5%	0%
Gr. 5 Math	N	2	19	0	14	4	1	2	0
	%	10%	90%	0%	67%	19%	5%	10%	0%
Gr. 6 Math	N	4	14	0	12	3	2	1	0
	%	22%	78%	0%	67%	17%	11%	6%	0%
Gr. 7 Math	N	4	16	0	14	4	2	0	0
	%	20%	80%	0%	70%	20%	10%	0%	0%
Gr. 8 Math	N	4	12	0	11	3		2	0
	%	25%	75%	0%	69%	19%	0%	13%	0%
Algebra 1	N	6	13	0	10	5	3	1	0
	%	32%	68%	0%	53%	26%	16%	5%	0%
Algebra 2	N	3	13	0	12	1	1	2	0
	%	19%	81%	0%	75%	6%	6%	13%	0%
Geometry	N	3	13	0	12	0	2	2	0
	%	19%	81%	0%	75%	0%	13%	13%	0%
Gr. 3 ELA	N	0	17	0	13	2	2	0	0
	%	0%	100%	0%	76%	12%	12%	0%	0%
Gr. 4 ELA	N	1	18	0	14	2	2	0	1
	%	5%	95%	0%	74%	11%	11%	0%	5%
Gr. 5 ELA	N	2	15	1	12	2	1	3	0
	%	11%	83%	6%	67%	11%	6%	17%	0%
Gr. 6 ELA	N	3	18	0	16	2	2	0	1
	%	14%	86%	0%	76%	10%	10%	0%	5%
Gr. 7 ELA	N	0	15	0	11	2	2	0	0
	%	0%	100%	0%	73%	13%	13%	0%	0%
Gr. 8 ELA	N	1	15	0	13	2	1	0	0
	%	6%	94%	0%	81%	13%	6%	0%	0%
Gr. 9 ELA	N	1	13	1	14	1	0	0	0
	%	7%	87%	7%	93%	7%	0%	0%	0%
Gr. 10 ELA	N	2	15	1	10	4	0	3	1
	%	11%	83%	6%	56%	22%	0%	17%	6%

Table 8: Educator Panel, Years Professional

Committee		Less than 1 year	1–5 years	6–10 years	11–15 years	16–20 years	21 or more years	NR
Gr. 3 Math	N	7	5	2	3	0	1	0
	%	39%	28%	11%	17%	0%	6%	0%
Gr. 4 Math	N	5	6	3	1	1	1	2
	%	26%	32%	16%	5%	5%	5%	11%
Gr. 5 Math	N	6	7	6	0	0	2	0
	%	29%	33%	29%	0%	0%	10%	0%
Gr. 6 Math	N	6	5	3	2	0	1	1
	%	33%	28%	17%	11%	0%	6%	6%
Gr. 7 Math	N	2	9	4	3	1	1	0
	%	10%	45%	20%	15%	5%	5%	0%
Gr. 8 Math	N	3	7	5	1	0	0	0
	%	19%	44%	31%	6%	0%	0%	0%
Algebra 1	N	5	9	2	1	0	1	1
	%	26%	47%	11%	5%	0%	5%	5%
Algebra 2	N	7	5	0	4	0	0	0
	%	44%	31%	0%	25%	0%	0%	0%
Geometry	N	3	4	7	0	0	2	0
	%	19%	25%	44%	0%	0%	13%	0%
Gr. 3 ELA	N	1	8	1	1	3	2	1
	%	6%	47%	6%	6%	18%	12%	6%
Gr. 4 ELA	N	3	8	3	2	1	2	0
	%	16%	42%	16%	11%	5%	11%	0%
Gr. 5 ELA	N	2	6	3	3	1	2	1
	%	11%	33%	17%	17%	6%	11%	6%
Gr. 6 ELA	N	2	13	3	1	2	0	0
	%	10%	62%	14%	5%	10%	0%	0%
Gr. 7 ELA	N	1	5	4	2	1	1	1
	%	7%	33%	27%	13%	7%	7%	7%
Gr. 8 ELA	N	7	6	0	1	1	1	0
	%	44%	38%	0%	6%	6%	6%	0%
Gr. 9 ELA	N	3	6	1	4	0	0	1
	%	20%	40%	7%	27%	0%	0%	7%
Gr. 10 ELA	N	3	5	5	3	1	0	1
	%	17%	28%	28%	17%	6%	0%	6%

Table 9: Educator Panel, Years Teaching Experience

Committee		Less than 1 year	1–5 years	6–10 years	11–15 years	16–20 years	21 or more years	NR
Gr. 3 Math	N	0	0	6	7	1	4	0
	%	0%	0%	33%	39%	6%	22%	0%
Gr. 4 Math	N	0	1	4	5	5	4	0
	%	0%	5%	21%	26%	26%	21%	0%
Gr. 5 Math	N	0	3	7	5	3	3	0
	%	0%	14%	33%	24%	14%	14%	0%
Gr. 6 Math	N	0	0	8	4	0	6	0
	%	0%	0%	44%	22%	0%	33%	0%
Gr. 7 Math	N	0	2	7	8	1	2	0
	%	0%	10%	35%	40%	5%	10%	0%
Gr. 8 Math	N	0	0	4	5	3	4	0
	%	0%	0%	25%	31%	19%	25%	0%
Algebra 1	N	0	0	6	4	4	5	0
	%	0%	0%	32%	21%	21%	26%	0%
Algebra 2	N	0	0	5	1	4	6	0
	%	0%	0%	31%	6%	25%	38%	0%
Geometry	N	0	0	2	5	3	6	0
	%	0%	0%	13%	31%	19%	38%	0%
Gr. 3 ELA	N	0	0	4	7	2	4	0
	%	0%	0%	24%	41%	12%	24%	0%
Gr. 4 ELA	N	0	2	6	4	2	5	0
	%	0%	11%	32%	21%	11%	26%	0%
Gr. 5 ELA	N	0	1	8	3	4	2	0
	%	0%	6%	44%	17%	22%	11%	0%
Gr. 6 ELA	N	0	3	4	7	5	2	0
	%	0%	14%	19%	33%	24%	10%	0%
Gr. 7 ELA	N	0	0	3	5	6	1	0
	%	0%	0%	20%	33%	40%	7%	0%
Gr. 8 ELA	N	0	1	5	5	1	4	0
	%	0%	6%	31%	31%	6%	25%	0%
Gr. 9 ELA	N	0	0	4	5	4	2	0
	%	0%	0%	27%	33%	27%	13%	0%
Gr. 10 ELA	N	0	1	6	7	0	4	0
	%	0%	6%	33%	39%	0%	22%	0%

Table 10: Educator Panel, Years Teaching Experience in Assigned Subject/Grade

Committee		Less than 1 year	1–5 years	6–10 years	11–15 years	16–20 years	21 or more years	NR
Gr. 3 Math	N	3	7	4	3	1	0	0
	%	17%	39%	22%	17%	6%	0%	0%
Gr. 4 Math	N	3	8	5	1	1	0	1
	%	16%	42%	26%	5%	5%	0%	5%
Gr. 5 Math	N	1	11	6	0	2	1	0
	%	5%	52%	29%	0%	10%	5%	0%
Gr. 6 Math	N	0	5	7	5	0	1	0
	%	0%	28%	39%	28%	0%	6%	0%
Gr. 7 Math	N	1	6	8	5	0	0	0
	%	5%	30%	40%	25%	0%	0%	0%
Gr. 8 Math	N	2	4	6	3	1	0	0
	%	13%	25%	38%	19%	6%	0%	0%
Algebra 1	N	1	2	7	4	3	2	0
	%	5%	11%	37%	21%	16%	11%	0%
Algebra 2	N	0	1	8	1	5	1	0
	%	0%	6%	50%	6%	31%	6%	0%
Geometry	N	0	1	9	2	4	0	0
	%	0%	6%	56%	13%	25%	0%	0%
Gr. 3 ELA	N	0	5	4	5	1	1	1
	%	0%	29%	24%	29%	6%	6%	6%
Gr. 4 ELA	N	0	12	2	0	3	2	0
	%	0%	63%	11%	0%	16%	11%	0%
Gr. 5 ELA	N	3	7	3	2	2	0	1
	%	17%	39%	17%	11%	11%	0%	6%
Gr. 6 ELA	N	2	9	3	4	1	0	2
	%	10%	43%	14%	19%	5%	0%	10%
Gr. 7 ELA	N	1	3	4	2	3	0	2
	%	7%	20%	27%	13%	20%	0%	13%
Gr. 8 ELA	N	1	4	6	3	1	1	0
	%	6%	25%	38%	19%	6%	6%	0%
Gr. 9 ELA	N	1	7	5	1	1	0	0
	%	7%	47%	33%	7%	7%	0%	0%
Gr. 10 ELA	N	2	8	5	2	0	1	0
	%	11%	44%	28%	11%	0%	6%	0%

Table 11: Educator Panel Type of Community

Committee	Urban		Suburban		Rural		NR	
	N	%	N	%	N	%	N	%
Gr. 3 Math	3	17%	10	56%	5	28%	0	0%
Gr. 4 Math	4	21%	8	42%	5	26%	2	11%
Gr. 5 Math	8	38%	6	29%	7	33%	0	0%
Gr. 6 Math	6	33%	4	22%	7	39%	1	6%
Gr. 7 Math	10	50%	3	15%	7	35%	0	0%
Gr. 8 Math	9	56%	4	25%	3	19%	0	0%
Algebra 1	7	37%	7	37%	5	26%	0	0%
Algebra 2	4	25%	7	44%	5	31%	0	0%
Geometry	4	25%	6	38%	5	31%	1	6%
Gr. 3 ELA	3	18%	9	53%	4	24%	1	6%
Gr. 4 ELA	8	42%	2	11%	8	42%	1	5%
Gr. 5 ELA	5	28%	7	39%	4	22%	2	11%
Gr. 6 ELA	5	24%	10	48%	5	24%	1	5%
Gr. 7 ELA	5	33%	5	33%	5	33%	0	0%
Gr. 8 ELA	6	38%	4	25%	5	31%	1	6%
Gr. 9 ELA	4	27%	7	47%	4	27%	0	0%
Gr. 10 ELA	5	28%	8	44%	4	22%	1	6%

Table 12: Educator Panel Region of the State

Committee	Panhandle		Northeast		East Central		West Central		South		NR	
	N	%	N	%	N	%	N	%	N	%	N	%
Gr. 3 Math	4	22%	3	17%	3	17%	3	17%	5	28%	0	0%
Gr. 4 Math	4	21%	4	21%	4	21%	4	21%	3	16%	0	0%
Gr. 5 Math	2	10%	3	14%	5	24%	6	29%	5	24%	0	0%
Gr. 6 Math	3	17%	3	17%	3	17%	2	11%	6	33%	1	6%
Gr. 7 Math	4	20%	4	20%	4	20%	3	15%	5	25%	0	0%
Gr. 8 Math	1	6%	1	6%	4	25%	3	19%	7	44%	0	0%
Algebra 1	3	16%	2	11%	3	16%	3	16%	8	42%	0	0%
Algebra 2	2	13%	1	6%	5	31%	3	19%	5	31%	0	0%
Geometry	3	19%	2	13%	3	19%	3	19%	4	25%	1	6%
Gr. 3 ELA	3	18%	2	12%	4	24%	4	24%	4	24%	0	0%
Gr. 4 ELA	3	16%	2	11%	5	26%	4	21%	5	26%	0	0%
Gr. 5 ELA	4	22%	3	17%	3	17%	4	22%	3	17%	1	6%
Gr. 6 ELA	3	14%	2	10%	7	33%	5	24%	4	19%	0	0%
Gr. 7 ELA	1	7%	4	27%	3	20%	3	20%	4	27%	0	0%
Gr. 8 ELA	3	19%	2	13%	5	31%	3	19%	3	19%	0	0%
Gr. 9 ELA	2	13%	4	27%	2	13%	3	20%	4	27%	0	0%
Gr. 10 ELA	5	28%	3	17%	2	11%	3	17%	5	28%	0	0%

Table 13: Educator Panel District Size

Committee	Large		Medium		Small		NR	
	N	%	N	%	N	%	N	%
Gr. 3 Math	8	44%	5	28%	5	28%	0	0%
Gr. 4 Math	7	37%	7	37%	4	21%	1	5%
Gr. 5 Math	9	43%	7	33%	5	24%	0	0%
Gr. 6 Math	7	39%	3	17%	8	44%	0	0%
Gr. 7 Math	9	45%	7	35%	4	20%	0	0%
Gr. 8 Math	9	56%	2	13%	5	31%	0	0%
Algebra 1	10	53%	5	26%	3	16%	1	5%
Algebra 2	9	56%	2	13%	5	31%	0	0%
Geometry	6	38%	3	19%	7	44%	0	0%
Gr. 3 ELA	8	47%	3	18%	5	29%	1	6%
Gr. 4 ELA	8	42%	3	16%	8	42%	0	0%
Gr. 5 ELA	7	39%	5	28%	5	28%	1	6%
Gr. 6 ELA	7	33%	9	43%	5	24%	0	0%
Gr. 7 ELA	7	47%	1	7%	5	33%	2	13%
Gr. 8 ELA	7	44%	5	31%	4	25%	0	0%
Gr. 9 ELA	5	33%	5	33%	5	33%	0	0%
Gr. 10 ELA	9	50%	8	44%	1	6%	0	0%

During the Educator Panel meeting, panelists were divided into 17 rooms, one room for each grade and subject. The discussions in each room were guided by AIR facilitators and assistants to the facilitators. The facilitator explained procedural constructs of the process, conducted training with the online tools, led discussions for four rounds of standard setting, maintained security of test content, collaboratively kept schedules consistent across all rooms, fielded questions from the panelists, and ensured that timely recommendations were provided to the AIR psychometricians for statistical calculations. The room assistant was responsible for providing assigned materials and ensuring the security of test materials at all times. The number of panelists for each subject and their assigned facilitators are presented in Table 14.

Table 14: Educator Panel Configuration

Panel	Panelists	Table Leaders	Total	Subject	Grade/ EOC	AIR Facilitator	AIR Facilitator Assistant
1	15	3	18	Mathematics	3	Alysa Kartee	Tiffany Abu-Shaikha
2	16	3	19	Mathematics	4	Jim McCann	Lisa Schaaf
3	17	4	21	Mathematics	5	Paul Maxon	Daniel Freedberg
4	15	3	18	Mathematics	6	Erica Ajder	Eileen Heneghan
5	16	4	20	Mathematics	7	Maureen Font	Nate Thompson
6	13	3	16	Mathematics	8	Jennifer Rubel	Bernard Farley
7	16	3	19	Algebra 1	EOC	Kari Stellpflug	Christina Estes
8	13	3	16	Algebra 2	EOC	Chris Paskoff	Susan Sherwood
9	13	3	16	Geometry	EOC	Sam Thomas	Marie Kristine-Tardif
10	14	3	17	ELA	3	Allison Stingley	Stephanie Ryan
11	16	3	19	ELA	4	John Neral	Jacob Wilkes
12	15	3	18	ELA	5	Sean Redmond	Kevin Clayton
13	17	4	21	ELA	6	Brett Craycraft	Sarah Abdelnaby
14	13	3	15	ELA	7	Diana Reed	Terra Winsett
15	13	3	16	ELA	8	Natalie Rebentisch	Amber Benlian
16	13	3	15	ELA	9	Kelly Quinney	Brian Kline
17	15	3	18	ELA	10	Katina Marshall	Anthony Kazanjian
						June Zack	Mathematics Lead
						Kevin Dwyer	ELA Lead
Total	250	54	302				

The FDOE and TDC staff in attendance were also active monitors of the Educator Panel meeting. They provided the panelists with additional information and answered panelists' questions on the assessment, content, policy, and historical aspects of assessment in general sessions, breakout rooms as needed, and individually. The participants who attended the meeting are listed in Table 15.

In addition to the meeting facilitators, a number of other staff from AIR attended the Educator Panel meeting. Table 16 lists these staff and their respective roles.

Table 15: Educator Panel Meeting, Attendees from FDOE and TDC

	Attendee	Affiliation	Role
1	Commissioner Pam Stewart	FDOE	Introductory Remarks
2	Juan Copa	FDOE	Introductory Remarks
3	Vince Verges	FDOE	Introductory Remarks
4	Victoria Ash	FDOE	Observer
5	Molly Hand	FDOE	Observer
6	Salih Binici	FDOE	Psychometric team
7	Mengyao Cui	FDOE	Psychometric team
8	Sanghyun Jeon	FDOE	Psychometric team
9	Zhongtian Lin	FDOE	Psychometric team
10	Steve Ash	TDC	Observer
11	Renn Edenfield	TDC	ELA Lead Observer
12	Sally Rhodes	TDC	ELA Observer
13	Elizabeth Tricquet	TDC	ELA Observer
14	Gretchen Sims	TDC	ELA Observer
15	Sally Donnelly	TDC	ELA Observer
16	Chris Harvey	TDC	Mathematics Lead Observer
17	Terri Sebring	TDC	Mathematics Observer
18	June Campbell	TDC	Mathematics Observer
19	Travis Barton	TDC	Mathematics Observer

Table 16: Other AIR Staff Members in Attendance

Group	Attendees
Overall Coordinator	<ul style="list-style-type: none"> ▪ Gary Phillips
Project Management staff	<ul style="list-style-type: none"> ▪ Kevin Murphy ▪ Evelyn Chester ▪ Liz Mortimer ▪ Meredith Durgin ▪ Hannah Binder
AIR Psychometric support	<ul style="list-style-type: none"> ▪ Harold Doran ▪ Tao Jiang ▪ Bokhee Yoon ▪ MinJeong Shin ▪ Dipendra Subedi
Overall AIR Statistical support	<ul style="list-style-type: none"> ▪ Nicholas Kalich ▪ Patrick Kozak ▪ Jessica Crutchfield ▪ Sydney Fitzgerald ▪ Hashim Evans ▪ Danielle Peterson ▪ Alex Mendoza
Backup facilitators	<ul style="list-style-type: none"> ▪ Kathy Laya ▪ Heidi Beeman ▪ Heather Williams ▪ Crystal Davidson
IT Support staff	<ul style="list-style-type: none"> ▪ Abdul-Hadi Sid Ahmed ▪ Eric Rose

During the Educator Panel meeting, three to four panelists per room served as table leaders for their respective rooms. On the first day of the meeting, special training was provided to these table leaders to prepare them appropriately. Table leaders were selected on the basis of knowledge of the process and experience in their fields. They were expected to have a broad perspective of the process and to assist in communication between AIR staff members and other panelists. Table leaders were tasked with assisting standard setting staff by

- facilitating discussions at their table;
- assisting with distribution and collection of standard setting meeting materials; and
- alerting meeting staff of confusion or concerns at their tables.

Throughout the standard setting process, panelists viewed live test items and other confidential assessment materials. Table leaders were asked to assist in ensuring that all secure materials

remained in the meeting rooms and that all cell phones were properly stored; any violation of the security affidavit was to be reported to facilitators.

3.2 LOGISTICAL PREPARATION

The standard setting meetings were held at the Grand Cypress Hotel in Orlando, Florida. AIR acquired 17 rooms in order to convene panels of educators for each individual grade and subject. The timeline for completing the standard setting meeting was reasonable as it was spread over five days for ELA, Mathematics, and EOC assessments. An outline of the agenda is provided in Appendix C.

The Educator Panel meeting utilized separate rooms for training, psychometric data analysis, AIR and FDOE staff meetings, workrooms, and a secure room for material storage and preparation. There was one large room with seating for about 400 individuals, used for large group orientation and presentations. Additionally, each of the 17 panel breakout rooms was arranged to accommodate three to four tables of panelists, with sufficient space for both a laptop and writing/working space for each participant. These meeting rooms were equipped with technological materials, such as LCD projectors and one laptop computer per panelist, with hard-wired Internet connection. The laptops were utilized to access the FSA Ordered Item Booklet, along with AIR's web-based systems utilized throughout the process. Appendix F describes the physical and online materials provided. Table 17 describes the room logistics, organization, and technology requirements for each of the meeting rooms.

Table 17: Summary of Required Meeting Space

Room Use	Room	Number Tables	Number per table	Type of Room	Observer table	Estimated Capacity Needed	IT requirement
Large group training	Large Group (all panelists)	--	--	Podium	--	420	Projector, presenter microphone
Table leader training	Table Leader ELA	18	5	Podium	--	90	Projector
Break out	ELA 3	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	ELA 4	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	ELA 5	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	ELA 6	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	ELA 7	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	ELA 8	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	ELA 9	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	ELA 10	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Mathematics 3	4	5	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Mathematics 4	3-4	5-7	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Mathematics 5	3-4	5-7	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Mathematics 6	3-4	5-7	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Mathematics 7	3-4	5-7	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Mathematics 8	3-4	5-7	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Algebra 1	3-4	5-7	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Geometry	3-4	5-7	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
	Algebra 2	3-4	5-7	Banquet	Space for 4 individuals	26	Projector; hard-wired Internet; flip chart
Data Analysis	Psycho-metrician	--	--	--	--	10	Projector; hard-wired Internet
Staff work space	AIR Staff	--	--	--	--	20	Wi-Fi
	FDOE Staff	--	--	--	--	20	Wi-Fi
	AIR & FDOE	--	--	--	--	20	Projector; Wi-Fi
Material Storage and Prep	Storage	10	0	--	--	0	none

Considering the widespread nature of the meeting, security was a crucial consideration of the Educator Panel meeting. In order to protect the validity of the assessment, it was critical to always maintain the security of the FSA items. In addition to having panelists sign non-disclosure agreements, which included Florida’s Test Security Statute, 1008.24, as well as State Board of Education Test Security Rule, 6A-10.042, F.A.C., AIR used additional security approaches, such as securing items with password-protected access and prohibiting the use of technology in the panel rooms. AIR also kept all physical data under tight security. For example, the data analysis workroom was kept locked and monitored by AIR staff at all times. As an added precaution, AIR staff constantly monitored entry into the participant workrooms as well as the project workroom, the data processing room, and the staff meeting room. Appendix D contains a complete discussion of all security measures.

3.3 TRAINING

Prior to the Educator Panel, it was necessary to ensure that each AIR facilitator was extensively knowledgeable of the intellectual constructs and technologies used in standard setting. Adequate training was also essential to standardize the training and procedures across the grade/subject committees. AIR staff members received extensive training for their respective roles in the process.

Training the panelists on the bookmarking method was essential for the Educator Panel meeting. Prior to any exposure of the secure materials, panelists were required to sign non-disclosure agreements; all panelists agreed and complied with this requirement. See Appendix E for details. The training process was led by the AIR meeting facilitators and involved a review and discussion of the Florida Standards, the test specifications, the ALDs for each achievement standard, the Ordered Item Booklet, and Response Probability. Furthermore, table-leader training emphasized the necessity of securing materials. FDOE reviewed and approved all training materials used at the Educator Panel prior to the meeting.

AIR room facilitators were assigned to each of the committees in the breakout rooms to provide training on the content, test specifications, and ALDs. They provided the panelists with materials on the content standards and test specifications as well as an explanation of how the FSA was developed from the content standards. Before setting any cut scores, panelists were instructed to familiarize themselves with the content standards, test designs, and intellectual expectations of students.

3.3.1 Taking the Test

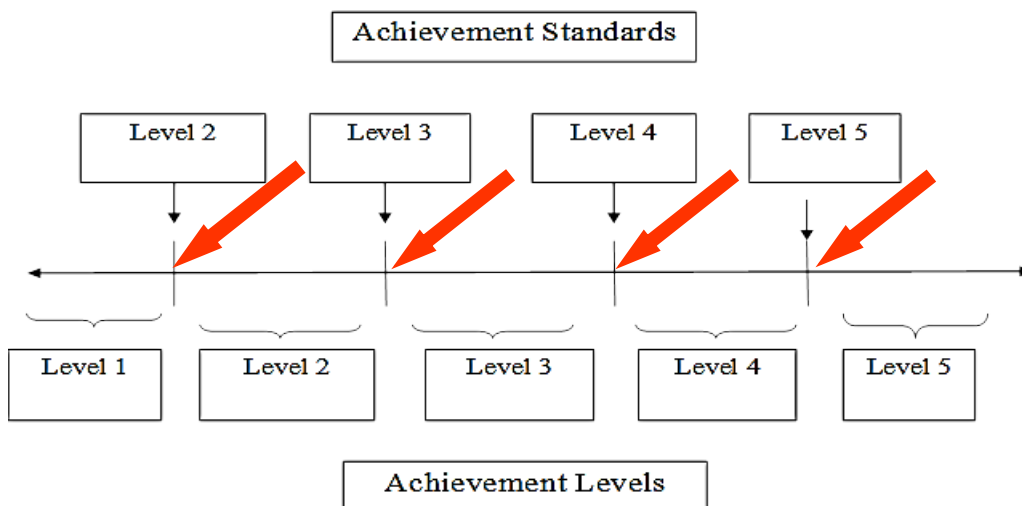
Panelists were then administered the FSA for their given grade and subject in the exact same testing environment that was administered to students. Presentations by the facilitators helped guide this process.

3.3.2 “Just Barely” Achievement Level Descriptions

After taking the test, panelists familiarized themselves with the Achievement Level Descriptions previously described in Section 2. They were then asked to form “Just Barely” Achievement Level Descriptions, which defined the expectations for lowest performing students in each

achievement level. A “Just Barely” template and an example were developed by AIR for each grade and subject prior to the meeting of the Educator Panel, and the panelists then used the template to develop their own “Just Barely” ALDs through discussion. These descriptions represented the minimum expectations associated with each achievement level. The red arrows in Figure 20 illustrate that the “Just Barely” ALDs defined the very lowest boundaries of the Range ALDs. These “Just Barely” descriptions helped to narrow the focus of the panelists to the most basic, essential knowledge and skills required to meet each achievement level. The “Just Barely” descriptions developed by the panelists are listed in Appendix B.

Figure 20: “Just Barely” Descriptions

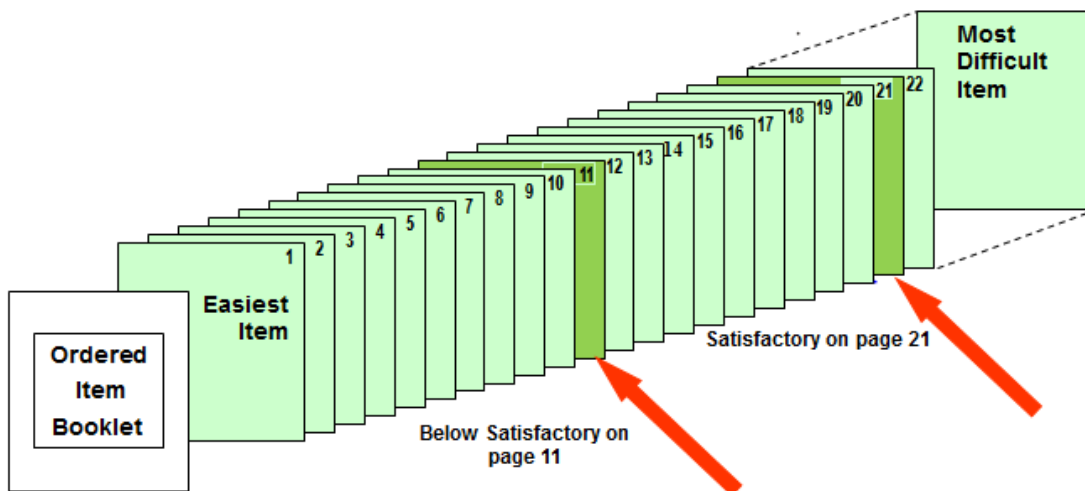


3.3.3 Ordered Item Booklet

The Bookmark method used in standard settings around the country utilizes ordered item booklets (OIBs) as a central tool for the setting of standards. OIBs are sets of test items, for every grade and subject, ordered by difficulty (not the order in which the items appear on actual test form). For the FSA Educator Panel, each OIB contained 50–60 FSA items, depending on the test blueprints. During the rounds of the Educator Panel, panelists placed “bookmarks” in these ordered test items to identify where the cut boundaries should be for each achievement level. Each page of the OIB corresponded to a scale score cut score; thus, when panelists placed their “bookmarks” for each achievement level, they were in fact selecting achievement standards for the achievement levels.

Figure 21 illustrates how the bookmark placement for ALDs is accomplished. In the figure, the items are ordered from easiest to hardest across the OIB. The panelists use the content standards and ALDs to locate the item that best describes the lowest bound of each achievement standard.

Figure 21: Bookmark Placement



FDOE suggested four rounds of this bookmark placement in order to allow for ample discussion among the educators and to gradually introduce external statistical information with each successive round to better inform their judgments.

In developing the OIBs, the decision was made to augment the operational forms with field test items, in order to smooth the frequency distribution of item difficulty. Increasing the number of items created a more even distribution of difficulty, thus providing panelists with greater context to identify important shifts in the knowledge and skill requirements of test items. With the Bookmark method, panelists may become focused on the cognitive demands of a single item when deliberating on the location of a particular performance standard, and this propensity is exacerbated when there are relatively few items in a given location. In addition, there are typically fewer items available in locations associated with more extreme standards, such that movement of the bookmark by even a page or two may result in very large increases or decreases in the percentage of students meeting the standard. Augmenting the OIB can significantly moderate the impact associated with each OIB page, even for more extreme cuts. All adjustments made to the OIB maintained the representativeness and the proportionality to the blueprint. FDOE and TDC monitored the OIB modifications and approved the final OIBs that were used during standard setting.

Items on the OIB were ranked according to their response probability (RP) value, which is based on Item Response Theory parameters. The RP criterion refers to the location in the OIB that corresponds to a given probability of success. RP value of 0.67 was used in FSA tests in grades 3 through 10 ELA and grades 3 through 8 Mathematics. The RP67 value indicates that the student who just barely reaches the standard has a 0.67 probability or likelihood of answering the item correctly in the case of a multiple choice item or a 0.67 probability of getting the corresponding score in the case of a constructed response item (Huynh, 2006). For constructed-response items, the ordering was based on step-level RP67 values for ELA grades 3–10 and Mathematics grades 3–8. Because the EOC assessments were substantially more difficult to the population tested, an

RP value of 0.50 was used for these tests only. Constructed-response items, including the writing prompts in Grades 4-10 ELA, appeared multiple times in the OIB, once for each score point.

Before setting their cut points for each round, educators were first asked to review the OIB and make content judgments about each item. Next, using the content-driven ALD and “Just Barely” ALD as references, they placed a bookmark beside the item that just barely qualified for the achievement levels. These judgments were based on their experience, knowledge of the content standards, training, and the given response probability (RP67) level.

The OIBs were presented to panelists electronically, on laptops provided by AIR, in AIR’s standard setting tool. This web-based standard setting tool allowed each panelist to interact with the item as it was administered and was also used to collect the actual bookmarks recorded by the panelists. The web-based standard setting tool not only presented the items of the OIB, but it also displayed the domain, cluster, and standard of each particular item, the correct answer, the score points, a section for notes, and a tab that displayed statistical feedback introduced throughout the process.

Note that the ELA Ordered Item Booklets for grades 4 through 10 contained items reflecting all score points associated with the Reading items as well as the score points associated with the Writing prompt. Writing responses were scored along three dimensions, each worth a different number of points: Conventions (2 item parameters), Elaboration (3 item parameters), and Organization (3 item parameters). Thus, the prompt contributed eight pages to the OIB. In the OIB, points for Conventions were labeled as 1 and 2. For Elaboration and Organization, the points were labeled 2, 3, and 4. Elaboration and Organization did not have a score point of 1 because these rubric categories had four points ranging from one to four, but there was no score point of zero. When the prompts were calibrated, we estimated the difficulty of the steps from 1 to 2, 2 to 3, and 3 to 4. Hence, without a zero point in the rubric, it was not possible to estimate a step for 0 to 1; therefore, this score point did not exist.

3.4 MARKING THE PAGES IN THE ORDERED ITEM BOOKLET

AIR’s standard setting web tool was used by each individual panelist to place bookmarks in the OIB, which represented recommended cut scores for each achievement level. Psychometricians analyzed the cut score recommendations from the panelists and provided graphical and statistical feedback throughout the process. Furthermore, AIR and FDOE psychometricians also participated in the meetings as necessary when panelists raised questions regarding statistical analyses during panelist discussions. There was one practice round followed by four rounds of standard setting, structured as described below.

3.4.1 Practice Round

Panelists used a seven-to-ten item OIB designed to give them an understanding of the bookmarking process and how to recommend a cut score using AIR’s online tool. These items reflected a range of item types and were used as a reference point for further discussion for the setting of cut points. The purpose of the Practice Round was to ensure that panelists were comfortable with the technology and item types prior to setting any actual bookmarks.

3.4.2 Round 1

Before setting any bookmarks, panelists completed their discussions of the ALDs, the Just Barely descriptions, and the OIB. They were then required to sign the Readiness Form (see Appendix I for an example), indicating that they understood the task at hand and were ready to make their cut score recommendations. Panelists were once again asked to consider characteristics of a student who would just barely represent each achievement level, and they then made independent judgments about the page in the OIB where the student would have about a two-thirds chance of getting the item correct; i.e., an RP value of 0.67. The panelists bookmarked the cut scores using AIR’s standard setting web tool. In each round, panelists made the Level 3 recommendation first, followed by Level 2, Level 4, and Level 5. They were instructed to allocate the majority of their time to the consideration of Level 3, which was intended to help anchor the remaining levels. While setting their bookmarks, it was also necessary that the bookmarks were ultimately sequential to reflect the ordered achievement levels outlined in Section 2.

Panelists received and discussed feedback from their Round 1 ratings for tables and the entire room. The feedback provided to the panels was in the form of median ratings of the OIB page numbers. An example of a feedback table from Round 1 is displayed in Table 18. Medians were used because page numbers represent ordinal, not interval data.

Table 18: Sample Feedback from Round 1

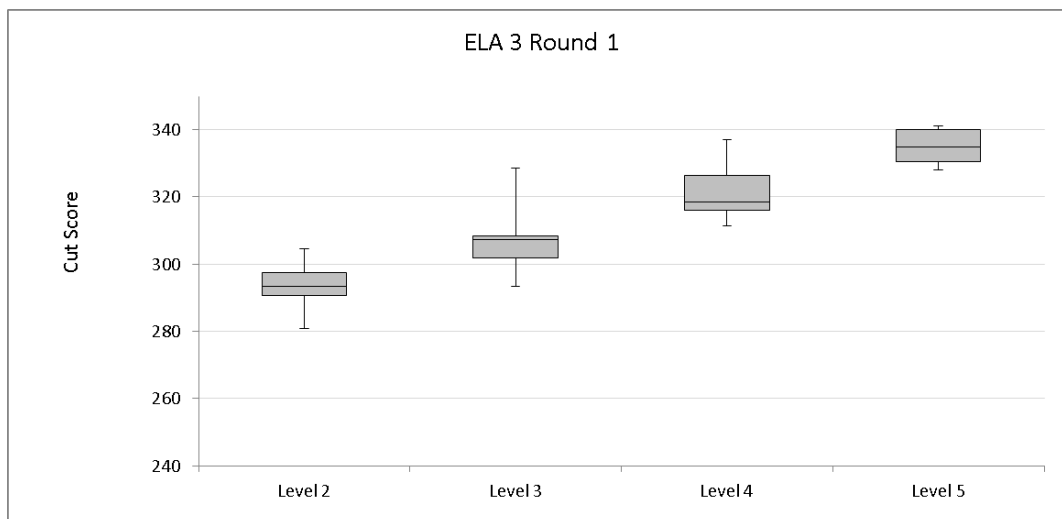
Table Report for ELA 3				
Table 1	Round 1 Pages in OIB			
	Level 2	Level 3	Level 4	Level 5
Panelist 1	11	26	36	48
Panelist 2	10	19	32	48
Panelist 3	11	24	40	49
Panelist 4	11	21	35	48
Panelist 5	6	11	29	45
Panelist 6	20	47	48	54
Table 1 Median	11.00	22.50	35.50	48.00

Room Report for ELA 3				
	Round 1 Pages in OIB			
	Level 2	Level 3	Level 4	Level 5
Table 1	11.00	22.50	35.50	48.00
Table 2	10.00	24.00	36.00	54.00
Table 3	10.50	23.50	39.00	51.50
Room Median	11	24	36	51

The variability in the panelists’ ratings was also evaluated using a Box and Whisker plot as illustrated in Figure 22. In this graph, the median scale score was used as a measure of central tendency, the first and third quartile were used as measures of dispersion, and the minimum and

maximum recommended scores were used for the whiskers. In order to observe the variability in panelists' ratings across rounds, there were four plots produced per grade/subject, one for each round. Figure 22 provides one example for ELA 3 from Round 1. Appendix O provides such plots for all grades and subjects.

Figure 22: Variability in Panelists' Ratings



In addition, the panelists received articulation feedback, which is further addressed in the Impact Data, Articulation, and Benchmarking section. The purpose of the articulation feedback was for the panelists to examine how their standards compared with the standards being recommended in other rooms. The articulated page numbers were given to the room facilitators, who then used that information to guide group discussion.

3.4.3 Round 2

Following further discussion considering both the “Just Barely” ALDs and the Round 1 feedback, each participant once again made an independent Round 2 judgment about the page in the OIB where the student has about a two-thirds chance of getting the item correct. Panelists again recommended Level 3 first, followed by Level 2, Level 4, and Level 5.

Panelists received and discussed graphical feedback provided through the standard setting tool from their Round 2 ratings for their individual tables and the entire room. As described above, the feedback consisted of statistics that described the central tendency and variability of the panelists' ratings, as well as articulation graphs to show the consistency of recommended standards across grades. Appendix P describes the results from Round 2.

3.4.4 Round 3

Prior to setting Round 3 bookmarks, panelists were presented with statewide impact data that showed the percentages of Florida students who would meet or exceed a cut score at any given page of the OIB. The impact data shown to the panelists displayed the overall percentages for the total population. Thus, panelists had access to an empirical, external reference about the impact of their standard setting recommendations. An illustration of the impact data is provided in the

Impact Data, Articulation, and Benchmarking section. Once again, following these discussions, each participant made an independent Round 3 judgment about the page in the OIB where the student should have about a two-thirds chance of getting the item correct. Appendix Q specifies the results from Round 3.

Statistical feedback was once again provided to the panelists via the standard setting tool, and each room discussed the results of this round in relation to the content standards and statistical Florida-specific impacts of the cuts, always with an understanding of articulation across grades.

3.4.5 Round 4

At this stage of the process, benchmark data was provided to panelists, displaying the page number in the OIB that corresponded to external national and international benchmarks, such as the National Assessment of Educational Progress (NAEP), Program for International Student Assessment (PISA), and Trends in International Mathematics and Science Study (TIMSS). These benchmarks gave panelists an additional empirical external reference about the impact of their standard setting recommendations. Then, for a final time, each panelist made an independent Round 4 judgment about the page in the OIB where the student would have about a two-thirds chance of getting the item correct. Appendix R details the results from Round 4.

3.5 EVALUATION OF EDUCATOR PANEL MEETING

After all activities were completed, the panelists completed online meeting evaluations independently, in which they were given the opportunity to describe and assess their experience participating in the Educator Panel standard setting using the Bookmark method. The meeting evaluation form appears in Appendix K. A summary of the results from the meeting evaluation form is available in Appendix V.

3.6 IMPACT DATA, ARTICULATION, AND BENCHMARKING

Impact Data, Articulation, and Benchmarking were all critical statistically-driven components of the Educator Panel, as described in the previous sections. These were gradually introduced during the meeting, always to assist or enhance the content-driven Bookmark method. They are additionally referenced in the next section describing the process of the Reactor Panel to determine adjustments of the cut scores.

3.6.1 Impact Data

The Educator Panel was presented with statewide impact data, based on actual data of Florida test takers. For this process, impact data is formally defined as the percentage of students meeting and exceeding any given achievement level for each page number in the Bookmark OIB. This data was utilized as a reference point for panelists to understand the implications of their content-based judgments. Estimations of impact data for each demographic group were based on observed scaled score distribution from the 2015 operational testing administration.

Calculating impact data requires estimating how well the students would have performed if they had been administered the representative form used during the standard setting. For RP theta, θ_0 ,

corresponding to each page number in the OIB, the scaled score, S_0 , is computed using the rounded value of $B + A\theta_0$, where B and A are the intercept and slope of theta to scaled score transformation equation presented in Table 19. For ELA and Mathematics, the transformation equation converts the on-grade theta into a vertically linked scale score. Section 6.4 in Volume 1 of 2015 Technical Reports provides an overview of the vertical scaling for ELA and Mathematics.

Table 19: Theta to Scale Score Transformation Equations

Subject	Grade	Theta to Scale Score Transformation
ELA	3	Scale Score= round(theta *20.000000 + 300.000000)
ELA	4	Scale Score = round(theta *20.237420 + 311.416960)
ELA	5	Scale Score = round(theta *21.230040 + 320.961420)
ELA	6	Scale Score = round(theta *21.861120 + 325.061500)
ELA	7	Scale Score = round(theta *21.581900 + 332.124320)
ELA	8	Scale Score = round(theta *21.531360 + 338.432720)
ELA	9	Scale Score = round(theta *21.751840 + 341.749740)
ELA	10	Scale Score = round(theta *21.284300 + 348.328540)
Mathematics	3	Scale Score= round(theta *20.000000 + 300.000000)
Mathematics	4	Scale Score = round(theta *20.899320 + 313.617800)
Mathematics	5	Scale Score = round(theta *22.050760 + 321.802560)
Mathematics	6	Scale Score = round(theta *21.684500+ 325.299220)
Mathematics	7	Scale Score = round(theta *20.379620 + 330.157540)
Mathematics	8	Scale Score = round(theta *19.952780 + 332.946420)
Algebra 1		Scale Score= round(theta *25.000000 + 500.000000)
Algebra 2		Scale Score= round(theta *25.000000 + 500.000000)
Geometry		Scale Score= round(theta *25.000000 + 500.000000)

Let S_i be the scaled score of the i th student who is eligible for state level score reporting, then the proportion of the population achieving the standard corresponding to the OIB page is the proportion of students whose scaled scores are at least S_0 , which is estimated by

$$\frac{1}{N} \sum_{i=1}^N 1_{S_i \geq S_0}$$

where N is the population n-counts, $1_{S_i \geq S_0}$ is defined as $1_{S_i \geq S_0} = \begin{cases} 1 & \text{if } S_i \geq S_0 \\ 0 & \text{if } S_i < S_0 \end{cases}$.

The same calculation is used to obtain impact data for each demographic group, in which case, the student population is defined as the eligible students from the corresponding demographic group.

Table 20 is an example of impact data that was available for the Educator Panel.

Table 20: Impact Data (Grade 3 ELA)

OIB Page	% Students	OIB Page	% Students	OIB Page	% Students	OIB Page	% Students
1	98.20	17	49.13	31	28.76	47	8.12
2	93.49	18	49.13	32	25.41	48	6.57
3	91.46	19	49.13	35	22.29	49	4.67
4	89.62	20	45.35	36	20.78	50	4.67
5	86.64	21	43.44	37	17.90	51	4.14
6	82.86	22	39.56	38	15.30	52	3.25
9	73.80	23	39.56	39	15.30	53	2.85
10	69.06	24	39.56	40	14.09	54	2.22
11	65.65	25	37.67	41	12.96	55	1.97
12	62.16	26	35.77	42	11.87	56	1.97
13	60.32	27	33.99	43	10.81	57	1.97
14	58.47	28	32.23	44	9.84	58	1.97
15	56.63	29	32.23	45	8.96		
16	54.78	30	30.52	46	8.96		

As an example, if a panelist were to select page 16 of the OIB for achievement level for Level 3, this standard would have been met by about 55% of the overall student population in this grade.

Impact data was presented to the Educator Panel for the following categories:

- Overall
- Gender
 - Male
 - Female
- Race/ethnicity
 - White
 - African American
 - Hispanic
 - American Indian
 - Asian
 - Pacific Islander
 - Multiracial

While the Educator Panel was presented with 10 demographic categories of impact data, the Reactor Panel was provided with impact data disaggregated by an even larger set of demographic categories as discussed in Section 4. Appendix T provides the impact data presented to the Educator Panel, and Appendix U presents the Reactor Panel impact data.

Exclusion Rules Used to Create Impact Data

When calculating impact data, certain exclusion rules were applied to meet the reporting specifications. A score status flag was available in the student data file for each test to facilitate the scoring and reporting process. The score status flag defined the reason whether a particular score was reported or not.

One of the key exclusion rules implemented excluded students with the score flag values other than 1. Each score status flag is defined below.

Score status 0: Not tested due to blank answer documents in paper-based testing (PBT) or for computer-based testing (CBT) any test that is started and closed without a single question being attempted

Score status 1: Score reported in regular reporting

Score status 2: Not attempted

Score status 3: Do not score (DNS) / Invalidations

Score status 4: Insufficient match to Test Information Distribution Engine (PBT only)

Score status 5: Below-grade tester for all tests except EOC

Score status 6: Duplicate record

Score status 7: FDOE holds, for reasons such as an invalid student ID and/or invalid date of birth

Score status 8: Caveon invalidated for potential test irregularity

Score status 9: Score reported in late reporting

Additionally, programs such as the Department of Juvenile Justice (DIJ) School, McKay Scholarship, Florida Corporate Tax Scholarship, Ahfachkee School, Private School, and Home Education were excluded according to the aggregate rules of the reporting specifications. The score flag rule covered the student-level exclusion rules such as demo students and the tests with *invalidated*, *expired*, and *reset* status, but did not remove out-of-grade-level test takers in non-EOC tests.

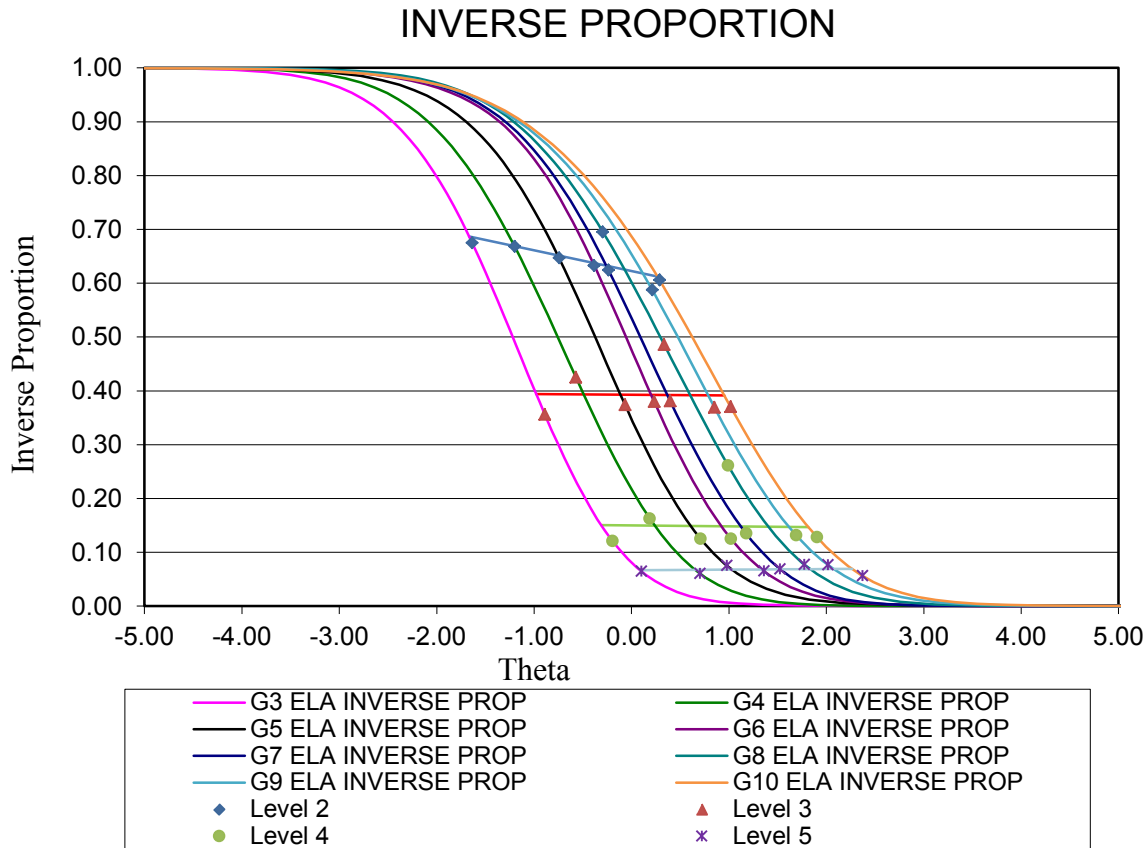
In grades 3 through 8 Mathematics, EOC, and grade 3 ELA, there was only one score flag, so any record flag value of other than 1 was excluded. However, in grades 4 through 10 ELA, there were two score flags corresponding to two components (Reading and Writing). Thus, only those records with both score flag values of 1 were used to create the impact data. Also, in all grades except grades 3 and 4, only online test takers were included in the population.

3.6.2 Articulation

When setting achievement standards, it is necessary to consider that the standards across the grades are to be reasonably consistent. It would not be logical, for example, to set high achievement standards in grade 3, low achievement standards in grade 4, and high achievement standards in grade 5 (Ferrara et al., 2007).

Figure 23 illustrates the concept of articulation using a hypothetical example of ELA data. After panelists recommended cut scores across the grades of ELA, a straight line was drawn for each achievement level from the lowest grade to the highest grade. In general, this represented the best-fitting regression line between the grades.

Figure 23: Articulation

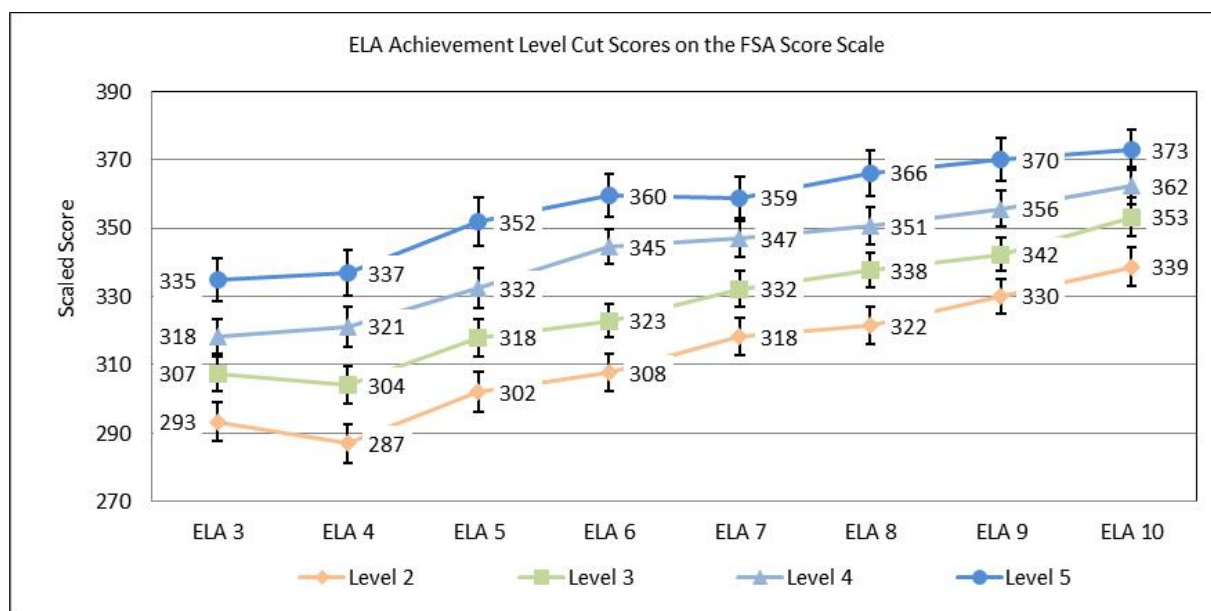


After Round 1 and Round 2, panelists were provided an example of a well-articulated set of standards. The articulated values of the page numbers served as a basis of discussion in the standard setting process. Educator Panelists were encouraged to consider these interpolated standards, but they were always reminded to use the content itself as the primary factor in their decisions. The extent to which the panelists adopted the interpolated achievement standards as their recommendations determined the overall articulation of the standards.

While impact data are normative, based on student data, and remain constant throughout the process, articulated information provided after each round to panelists would of course shift to reflect the bookmark medians after each round. The articulation information would tell the panelists what an articulated standard might be for the grade under consideration given the cut scores already recommended in the previous round and given the requisite content-referenced interpretations.

During the Educator Panel meeting, AIR’s standard setting web tool was also used to provide panelists with feedback following each round of standard setting. Figure 24 shows one example of feedback given. The error bars are the total standard errors of measurement ($SE_{Overall}$), which are composed of the conditional standard error of measurement ($CSEM$) at the cut score and the standard error of the cut score (SE_{Cut}), which is the standard error of the mean of the cut scores recommended by the panelists.

Figure 24: Round 1 Feedback to Panelists



Understanding the variability of cut scores among panelists, small groups, and rounds was of great interest. For each cut score, the variance associated with each of these sources was estimated using the restricted maximum likelihood estimate approach of the linear mixed model (Harville, 1977). For this study, the number of rounds was treated as a fixed factor, meaning that if the standard setting meeting was held again, the same number of rounds would be used. In addition, because panelists discussed all activities in small groups, their judgments were considered dependent on group membership. Therefore, panelists were considered *nested* within tables. Variance components for tables (σ_{Tables}^2) and panelists within tables ($\sigma_{Panelist Tables}^2$) were computed. Suppose there were N_{Tables} tables, and for the i th table, there were $N_{i,Panelist}$ panelists; then computation of the standard errors utilizes the following formula (Lee & Lewis, 2001):

$$SE_{cut} = \sqrt{\frac{\sigma_{Tables}^2}{N_{Tables}} + \frac{\sigma_{Panelist Tables}^2}{\sum_{i=1}^{N_{Tables}} N_{i,Panelist}} + \frac{\sigma_{Error}^2}{N_{Rounds} \sum_{i=1}^{N_{Tables}} N_{i,Panelist}}}$$

When median was used in standard setting, the above standard error was adjusted by $\sqrt{\frac{\pi}{2}}$.

Because the number of rounds was treated as a fixed facet, its variance component was not included in the error term. σ_{Error}^2 is a confounding term and includes the variance from the interaction between tables and panelists within tables, as well as variances unexplained by the defined facets. Standard errors were computed for each of the recommended cut scores.

The conditional standard error of measurement (CSEM) for each recommended cut score was calculated using the following formula:

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

where $I(cut)$ is the information of the cut score based on the standard setting form defined as

$$I(\theta) = \sum_{i \in MC} D^2 a_i^2 \frac{Q_i}{P_i} \left[\frac{P_i - c_i}{1 - c_i} \right]^2 + \sum_{i \in CR} D^2 a_i^2 \left(\frac{\sum_{m=1}^{K_i} m^2 \text{Exp}(\sum_{j=1}^m 1.7 a_j (\theta - b_{i,j}))}{1 + \sum_{m=1}^{K_i} \text{Exp}(\sum_{j=1}^m 1.7 a_j (\theta - b_{i,j}))} - \left(\frac{\sum_{m=1}^{K_i} m \text{Exp}(\sum_{j=1}^m 1.7 a_j (\theta - b_{i,j}))}{1 + \sum_{m=1}^{K_i} \text{Exp}(\sum_{j=1}^m 1.7 a_j (\theta - b_{i,j}))} \right)^2 \right)$$

where $D = 1.7$, $P_i = c_i + \frac{1 - c_i}{1 + \text{Exp}[-D a_i (\theta - b_i)]}$, $Q_i = 1 - P_i$, K_i is the maximum score point of the i th CR item starting from score 0.

The standard error of the cut score (SE_{cut}) and the conditional standard error of measurement (CSEM) were used to compute the overall standard error of the cut based on

$$SE_{overall} = \sqrt{SE_{cut}^2 + CSEM^2}$$

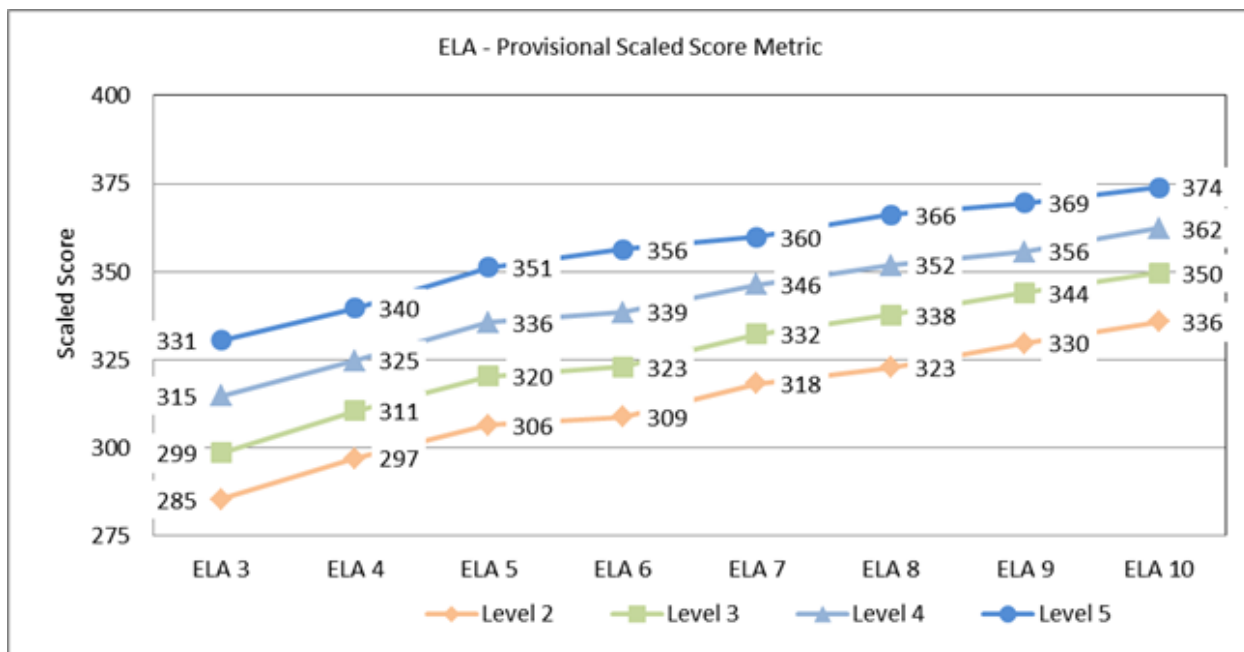
The $SE_{overall}$ is on the theta scale which can be transformed onto the scaled score scale by using $A \times SE_{overall}$, where A is the slope of theta to scaled score transformation equation defined in Table 19 earlier.

After each round, panelists would see a variety of graphical feedback displaying articulation information. As an example, Figure 24 represents the results of Round 1 judgments for ELA. The

panelists were shown the graph, and it was explained that this disjointed pattern of standards is unfavorable, as the Level 3 standard is higher in Grade 3 than in Grade 4.

The panelists were then shown a graph of articulated standards, such as the one presented in Figure 25. During the Educator Panel meeting, the OIB pages associated with these articulated standards were communicated to the panelists through the room leaders. In the following round, panelists were asked to discuss if there could potentially be content justifications for recommending standards in the vicinity of those provided in the articulated graph.

Figure 25: Articulation Feedback Given to Panelists



3.6.3 Benchmark Data

In addition to having well-articulated achievement standards across grades and subjects, FDOE wanted panelists to consider their recommendations when compared to national and international benchmarks as applicable. Benchmarking information was presented to the Educator Panel in Round 4 only; however, benchmark data were available to the Reactor Panel throughout their meeting. These data were presented to contribute to the discussions among panelists, and to use in making their Round 4 judgments. Florida used the approach recommended by AIR, which is outlined by Phillips (2011), in which the achievement standards are benchmarked against an external national and/or international referent, such as the NAEP, TIMSS, or PISA. AIR has used similar procedures in Oregon, Hawaii, Delaware, Minnesota, and Utah.

The most recent data from the NAEP was used, as presented in Table 21. Note that grade 12 NAEP was not appropriate as a benchmark for the FSA because the FSA ELA was administered in grades 3 through 10 and the FSA Mathematics was administered in grades 3 through 8.

Table 21: Percent in Florida at Each NAEP Standard

2013 NAEP	% at and Above
Reading Grade 4 Advanced	9%
Reading Grade 4 Proficient	30%
Reading Grade 4 Basic	36%
Reading Grade 4 Below Basic	25%
<hr/>	
Reading Grade 8 Advanced	3%
Reading Grade 8 Proficient	30%
Reading Grade 8 Basic	44%
Reading Grade 8 Below Basic	23%
<hr/>	
Mathematics Grade 4 Advanced	6%
Mathematics Grade 4 Proficient	35%
Mathematics Grade 4 Basic	43%
Mathematics Grade 4 Below Basic	16%
<hr/>	
Mathematics Grade 8 Advanced	7%
Mathematics Grade 8 Proficient	24%
Mathematics Grade 8 Basic	39%
Mathematics Grade 8 Below Basic	30%

The most recent international benchmark, TIMSS, was also used for Mathematics. Florida participated in the 2011 TIMSS assessment, so state-level representative data were available. Those results are in Table 22.

Table 22: Percent in Florida at each TIMSS International Benchmark

2011 TIMSS	Percent
Mathematics Grade 4 Advanced	14%
Mathematics Grade 4 High	33%
Mathematics Grade 4 Intermediate	36%
Mathematics Grade 4 Low	17%
<hr/>	
Mathematics Grade 8 Advanced	8%
Mathematics Grade 8 High	23%
Mathematics Grade 8 Intermediate	37%
Mathematics Grade 8 Low	32%

The 2012 PISA was also used for benchmarking. PISA is an age-based sample (age 15), not a grade-based sample. PISA results for Florida are presented in Table 23.

Table 23: Percent in Florida Reaching Each PISA Level

2012 PISA	Percent
Reading Level 6	0%
Reading Level 5	5%
Reading Level 4	20%
Reading Level 3	31%
Reading Level 2	26%
Reading Level 1	18%
Mathematics Level 6	1%
Mathematics Level 5	5%
Mathematics Level 4	13%
Mathematics Level 3	23%
Mathematics Level 2	28%
Mathematics Level 1	30%

The 2011 FCAT 2.0 (along with historical trends from 2011–2014) was also used in benchmarking, but this information was provided only to the Reactor Panel discussed in the next section, not the Educator Panel, which has been previously discussed.

Achievement Level Descriptions for NAEP, TIMSS, and PISA are contained in Appendix M. These were used by the panelists to familiarize themselves with the differences and similarities between the achievement levels on the FSA compared to those of NAEP, TIMSS, and PISA.

4. REACTOR PANEL MEETING

A Reactor Panel meeting convened for two days, September 10 and 11, 2015, to review the cut scores generated by the Educator Panel and make additional recommendations. This panel consisted of 16 Florida stakeholders (community leaders, education organization leaders, state university leaders, business leaders, school board members, and superintendents) to review and react to the Educator Panel’s cut score recommendations. The Reactor Panel was asked to review and modify the proposed cut scores, if they decided this was necessary. While the Educator Panel primarily made content-based judgments, the Reactor Panel was asked to focus on the impact of the proposed cut scores using impact data based on 2015 student performance, as well as data from external benchmarks and prior assessments (e.g., the historical trend for FCAT 2.0). The Commissioner of Education selected the participants for the Reactor Panel, and basic demographic data was collected using the form in Appendix H. This demographic information is presented in Table 24 and Table 25.

Table 24: Reactor Panel Attendees

	Name	Position	Representation	County
1	Ms. Marcia Andrews	School Board Member	School Board Member	Palm Beach
2	Dr. Nathan Balasubramanian	Executive Director, Strategy & Continuous Improvement	District	Broward
3	Mr. John Barnhill	Assistant Vice President for Enrollment Management at FSU	Postsecondary	Leon
4	Dr. Ed Bonahue	Provost and VP of Academic Affairs, Santa Fe College	Postsecondary	Alachua
5	Ms. Ronda Bourn	Assistant Director of Instruction for School Programs	District	NEFEC
6	Dr. Manoj Chopra	UCF Engineering Professor, former BOG member	Postsecondary	Orange
7	Ms. Karen Denbroeder	Retired Research Assistant, FL Center for Reading Research at FSU	SWD Expert	Leon
8	Mr. Robert Edwards	Superintendent	Superintendents	Lafayette
9	Mr. Danny Gaekwad	EFI Board of Directors	Business/Community	Marion
10	Dr. Diana Greene	Superintendent	Superintendents	Manatee
11	Dr. Barbara Jenkins	Superintendent	Superintendents	Orange
12	Debra Morton	Volunteer Coordinator at Fruit Cove Middle School	Parent	St. Johns
13	Dr. Susan Neimand	Director of School of Education, Miami Dade College	Postsecondary	Dade
14	Mr. Terry Nichols	School Board Member	School Board Member	Jackson
15	Rev. Ron Rawls	Pastor, St. Paul AME Church	Business/Community	St. Johns
16	Dr. Maria Torres	Executive Director of Federal and State Grants & English Language Learner Programs	ELL Expert	Collier

Table 25: Reactor Panel Demographics

Demographic	Level of Demographic	N	Aggregate Percentages
Gender	Male	8	50%
	Female	8	50%
Race/Ethnicity	White	8	50%
	African American	4	25%
	Hispanic	1	6%
	Native American	0	0%
	Asian	3	19%
	Other	0	0%
Location of Place of Employment	Urban	6	38%
	Suburban	4	25%
	Rural	3	19%
	No Response	3	19%
Geographic region	Panhandle	3	19%
	Northeast	4	25%
	East Central	2	13%
	West Central	1	6%
	South	3	19%
	No Response	3	19%

The Reactor Panel conducted its work in two rounds of judgment. In the first round, the Reactor Panel discussed the variation in cut scores recommended by the Educator Panel and provided independent ratings for any modifications to the cut scores. Appendix J provides the cut score recording forms. In the second round, the Reactor Panel reviewed the average cut scores from its Round 1 recommendations and was given an opportunity to modify any changes to the cut scores. However, of the 16 Reactor Panelists, no one modified his or her cut score recommendations in Round 2.

During the Reactor Panel’s review of the cut scores recommended by the Educator Panel, panelists were presented with impact data for the overall population plus a number of demographic subgroups displayed in a variety of graphical formats.

The Reactor Panel received the following types of FSA impact data graphs:

- Graph type 1 (2 line graphs and one point graph showing scale scores for achievement standards)
- Graph type 2 (3 bar charts for ELA, Mathematics, and EOCs)
- Graph type 3 (3 stacked bar charts within subjects, across grades & EOCs)
 - ELA grades 3–10
 - Mathematics grades 3–8
 - Algebra 1, Algebra 2, and Geometry

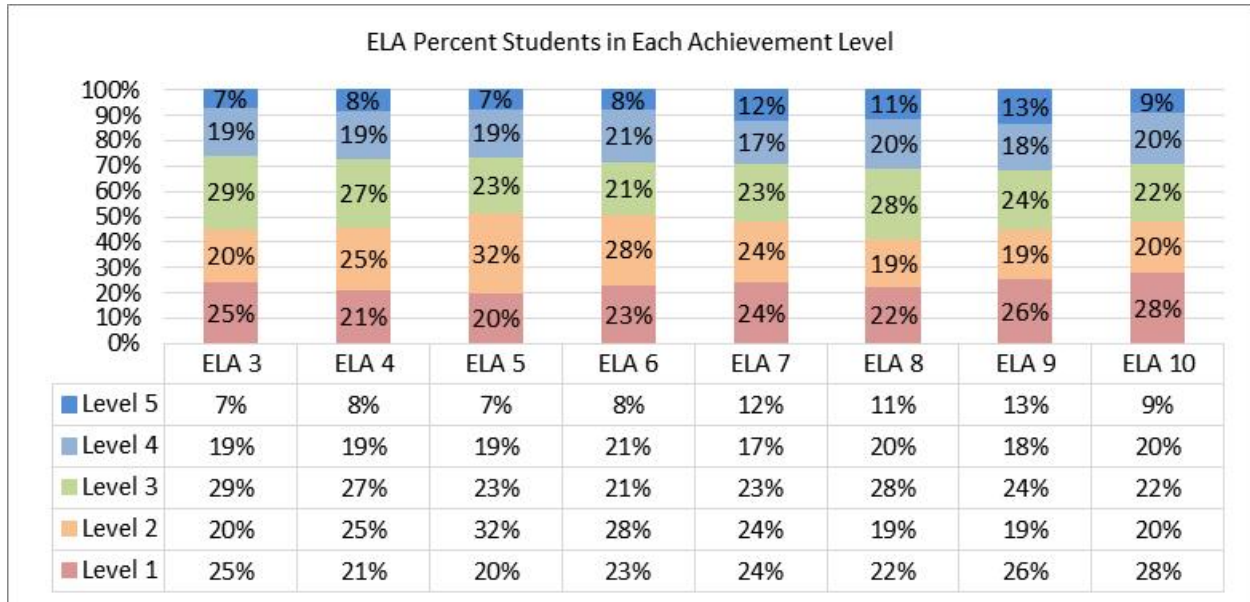
- Graph type 4 (17 stacked bar charts per subject/grade & EOC)
 - Overall
 - Male
 - Female
 - White
 - African American
 - Hispanic
 - American Indian
 - Asian
 - Pacific Islander
 - Multiracial
- Graph type 5 (17 stacked bar charts per subject/grade & EOC)
 - White Male
 - White Female
 - African American Male
 - African American Female
 - Hispanic Male
 - Hispanic Female
 - American Indian Male
 - American Indian Female
 - Asian Male
 - Asian Female
 - Pacific Islander Male
 - Pacific Islander Female
 - Multiracial Male
 - Multiracial Female
 - ELL
 - Non ELL
 - SWD
 - Non SWD

In addition, the Reactor Panel was given benchmark data that related Florida achievement to NAEP, TIMSS, and PISA. The Reactor Panel was also shown past FCAT 2.0 data from 2011–2014. This provided the Reactor Panel with information about the historical trend in Florida student achievement. Appendix N presents the FCAT 2.0 data provided to the Reactor Panel.

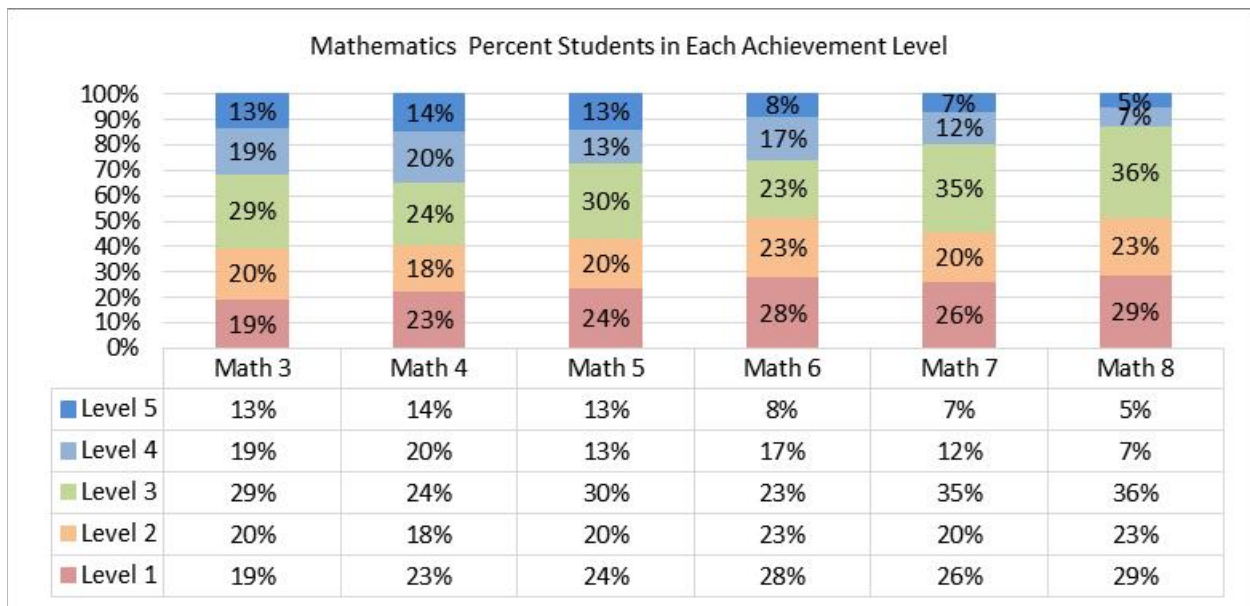
After panelists extensively discussed the purpose of their panel and the impacts of the given cut scores with the Commissioner and members of AIR and FDOE, they were given the opportunity to adjust the cut scores of the Educator Panel. As the Reactor Panel suggested cut score

revisions, the graphics were updated to reflect the new percentages associated with the revised cut score(s). Figure 26 through Figure 28 reflect what was shown to the Reactor Panel at the beginning of the meeting. Note that these results are based on the final round of the Educator Panel meeting.

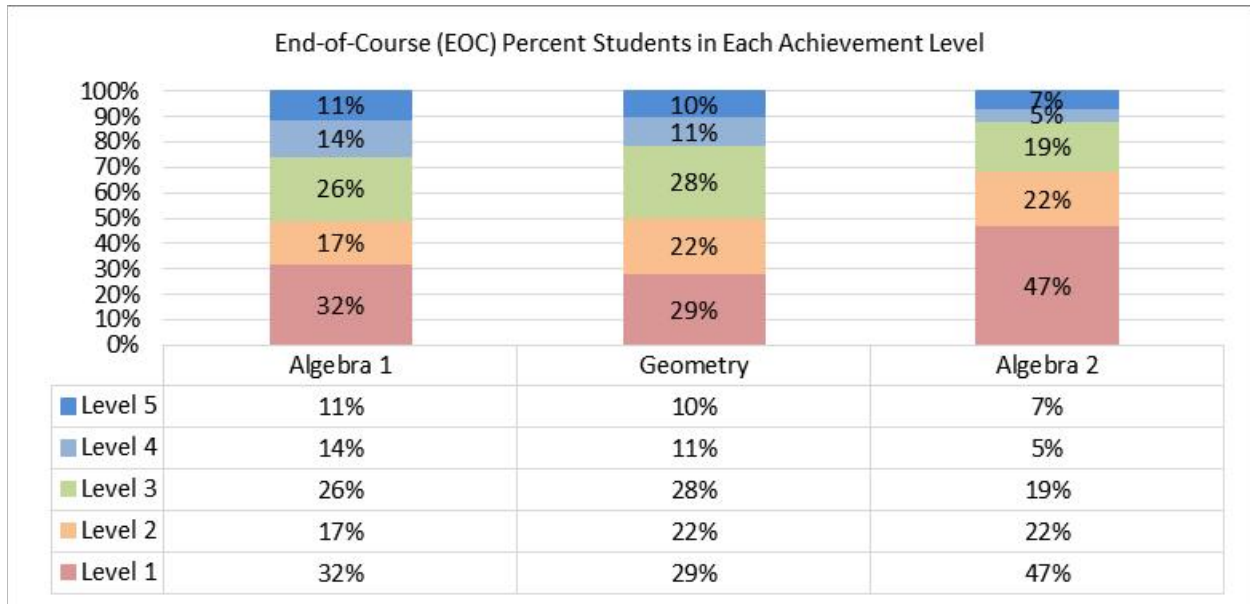
**Figure 26: Impact Data for Reactor Panel:
Percent Students in each Achievement Level for ELA**



**Figure 27: Impact Data for Reactor Panel:
Percent Students in each Achievement Level for Mathematics**



**Figure 28: Impact Data for Reactor Panel:
Percent Students in each Achievement Level for EOC**



Although the panelists were given the opportunity to adjust the cut scores in the second round, each panelist individually maintained his or her original recommendation. The summary of the Reactor Panel is presented in the executive summary section, and more detailed results from the Reactor Panel are presented in Appendix S. Finally, the panelists provided their evaluation of the Reactor Panel meeting to give feedback on the process; see Appendix L for the evaluation form used. A summary of the results from the evaluation form is available in Appendix W.

5. PUBLIC INPUT WORKSHOPS

After the Educator Panel and Reactor Panel meetings, public input on both panels' recommended cut scores was collected via public Rule Development workshops. These workshops were conducted in Ft. Lauderdale, Orlando, and Tallahassee and were also viewable via webcast. In his presentation outlining the standard setting process, Assistant Deputy Commissioner Vince Verges also provided the Educator Panel's and Reactor Panel's cut score recommendations and impact and benchmark data. After the presentation, questions were taken, and audience members were encouraged to submit feedback on feedback forms provided. An online survey was then provided on FDOE's website along with the workshop presentation itself. Additionally, individuals who were unable to attend one of the public input workshops had the opportunity to submit input for consideration via email, traditional mail, by fax, or by phone.

6. FINALIZATION OF ACHIEVEMENT STANDARDS

After the Educator Panel, Reactor Panel, and Rule Development workshops, the Department compiled feedback received through the Florida Administrative Register (FAR), the Office of Assessment (Assessment@fldoe.org), and the online feedback form posted on the Department website.

Taking into consideration the recommendations of both panels as well as public feedback, Commissioner Stewart recommended cut scores to reflect the state's expectations of student achievement. The Commissioner's recommended scores varied by no more than ± 4 score points from either the Educator Panel or Reactor Panel, and this was well within the range of variability of either or both of the panelists' judgment, depending on the grade and subject. In many cases, the recommendations of the Educator Panel, Reactor Panel, and Commissioner were the same.

In October, the Commissioner's recommended cut scores and the Rule development materials were provided to the legislature for the mandatory 90-day review period. The cut score recommendations were discussed at the December 2015 State Board of Education meeting, where board members requested additional information and clarification regarding the difference between Level 3 and Level 4 expectations. The Commissioner clarified for the State Board that level 3 is satisfactory (passing), while level 4 may be compared to proficient (higher performing), and that passing or satisfactory should be distinguished from proficiency in reported results.¹ To that end, FSA student reports were updated to provide additional detail to assist students, parents/guardians, and educators with interpreting student results in a meaningful way. More information about the revised student reports for use in 2016 and beyond can be found at: <http://www.fldoe.org/accountability/assessments/fsa-report.stml>.

¹ In an independent study comparing all states' performance standards to NAEP performance levels, which was completed after FSA standards were set, Gary Philips, Vice President for Psychometrics at AIR, concluded Florida's achievement level 4 aligns perfectly with NAEP Proficient. Refer to Volume 7 of this report for more information regarding the comparison of Florida's achievement levels to NAEP performance levels.

7. THE STANDARD SETTING RESULTS

This section outlines the results of the standard setting for ELA, Mathematics, and EOC assessments.

The results of achievement level cuts from the Educator Panel are presented in Table 26 through Table 28. Similarly, the results of achievement level cuts from the Reactor Panel are presented in Table 29 through Table 31. Finally, the recommendations from the Commissioner of Education are summarized in Table 32 through Table 34.

Table 26: Achievement Level Cuts in ELA from the Educator Panel

Grade	Level 2	Level 3	Level 4	Level 5
3	287	299	315	331
4	297	311	325	340
5	304	323	336	352
6	309	327	339	356
7	318	333	346	359
8	322	335	351	366
9	330	342	355	368
10	338	350	362	378

Table 27: Achievement Level Cuts in Mathematics from the Educator Panel

Grade	Level 2	Level 3	Level 4	Level 5
3	283	296	311	324
4	299	310	323	337
5	307	319	336	347
6	312	326	340	356
7	317	329	350	363
8	322	335	358	368

Table 28: Achievement Level Cuts in EOC Assessments from the Educator Panel

EOC	Level 2	Level 3	Level 4	Level 5
Algebra 1	488	500	518	532
Algebra 2	499	514	530	537
Geometry	486	501	521	533

Table 29: Achievement Level Cuts in ELA from the Reactor Panel

Subject	Grade	Level 2	Level 3	Level 4	Level 5
ELA	3	285	300	315	331
	4	297	310	325	340
	5	304	319	336	351
	6	309	325	339	356
	7	318	333	346	360
	8	322	337	352	365
	9	330	342	355	370
	10	338	350	362	378

Table 30: Achievement Level Cuts in Mathematics from the Reactor Panel

Subject	Grade	Level 2	Level 3	Level 4	Level 5
Mathematics	3	285	296	311	324
	4	299	309	323	337
	5	307	318	334	347
	6	312	325	341	356
	7	317	330	348	363
	8	322	337	356	368

Table 31: Achievement Level Cuts in EOC Assessments from the Reactor Panel

EOC	Level 2	Level 3	Level 4	Level 5
Algebra 1	487	494	518	532
Algebra 2	497	509	529	537
Geometry	486	497	521	533

Table 32: Achievement Level Cuts in ELA from the Commissioner

Subject	Grade	Level 2	Level 3	Level 4	Level 5
ELA	3	285	300	315	330
	4	297	311	325	340
	5	304	321	336	352
	6	309	326	339	356
	7	318	333	346	360
	8	322	337	352	366
	9	328	343	355	370
	10	334	350	362	378

Table 33: Achievement Level Cuts in Mathematics from the Commissioner

Subject	Grade	Level 2	Level 3	Level 4	Level 5
Mathematics	3	285	297	311	327
	4	299	310	325	340
	5	306	320	334	350
	6	310	325	339	356
	7	316	330	346	360
	8	322	337	353	365

Table 34: Achievement Level Cuts in EOC Assessments from the Commissioner

EOC	Level 2	Level 3	Level 4	Level 5
Algebra 1	487	497	518	532
Algebra 2	497	511	529	537
Geometry	486	499	521	533

On January 6, 2016, the State Board of Education voted (with six in favor and one opposed) to adopt the Commissioner’s recommended cut scores. Per state statute, students who took the Grade 10 ELA, Algebra 1, and Geometry assessments prior to the adoption of these cut scores are eligible to use the adopted Alternate Passing Scores indicated in Rule to meet their graduation requirements or Scholar designation requirements. The final language of the Rule is now available ([Rule 6A-1.09422, Florida Administrative Code](#)).

8. REFERENCES

- Ferrara, S., Phillips, G., Williams, P., Leinwand, S., Mahoney, S., & Ahadi, S. (2007). Vertically articulated performance standards: An exploratory study of inferences about achievement and growth. In R. Lissitz (Ed.), *Assessing and modeling cognitive development in school: Intellectual growth and standard setting* (pp. 31–63). Maple Grove, MN: JAM Press.
- Harville, D. A. (1977). Maximum-likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association*, 72(358), 320–340.
- Huynh, H. (2006). A clarification on the response probability criterion RP67 for standard settings based on Bookmark and item mapping. *Educational Measurement: Issues and Practice*, 25: 19–20.
- Lee, G., & Lewis, D. (2001, April). *A generalizability theory approach toward estimating standard errors of cutscores set using the Bookmark standard setting procedure*. Paper presented at the annual meeting of the National Council on Measurement in Education, Seattle, WA.
- 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* (pp. 249–282). Mahwah, NJ: Erlbaum.
- Phillips, G. W. (2011). The Benchmark method of standard setting. In G. J. Cizek (Ed.), *Setting performance standards: Foundations, methods, and innovations* (2nd edition). New York: Routledge.