APPENDIX B: STANDARDS AND STANDARD SETTING

- Executive Summary of the 2015 Georgia Milestones Standard Setting Technical Report
- Executive Summary of the 2018 Georgia Milestones Standard Setting Technical Report

Executive Summary of the 2015 Georgia Milestones Standard Setting Technical Report

Executive Summary

The Georgia Department of Education (GaDOE) partnered with Data Recognition Corporation (DRC) to conduct a standard setting workshop the Georgia Milestones Assessment System. Standard setting was conducted for grades 3–8 and end of course (EOC) in English language arts, mathematics, reading, science, and social studies. Committees of Georgia educators and stakeholders worked individually and in concert to recommend achievement standards associated with four achievement levels: *Beginning Learner*, *Developing Learner*, *Proficient Learner*, and *Distinguished Learner*. For reading, a single cut score divided students into two levels: Not *On-Grade Reading* and *On-Grade Reading.*

A three-phased process was conducted to establish achievement standards for the Georgia Milestones assessments:

1) Achievement level descriptors (ALDs) were developed by Georgia educators, and then edited by DRC and by the GaDOE. ALDs summarize the knowledge, skills, and abilities expected of students in each achievement level. Initial draft ALDs were developed in a workshop on March 23–24, 2015, and the drafts were later amplified in a second workshop on July 21–24, 2015.

2) Standard setting was conducted on August 24–29, 2015. The Bookmark Standard Setting Procedure (BSSP; Lewis, Mitzel, & Green, 1996; Lewis, Mitzel, Mercado, & Schulz, 2012) was implemented to recommend cut scores for the Georgia Milestones assessments. A total of 271 participants from across the state of Georgia participated in the standard setting workshop.

3) A review committee convened to consider the recommended achievement standards on September 1, 2015. A total of 22 Georgia educators, decision makers, and stakeholders considered the recommendations and made adjustments to promote articulation and consistency.

Each workshop was facilitated by researchers and content experts from DRC. GaDOE considered the recommended achievement standards in September 2015. After deliberation and presentation to the State Board of Education, GaDOE implemented the Board-approved cut scores. The Board approved the cut scores recommended by the standard setting and review committees on September 3, 2015. In December 2015, cut scores for new EOC assessments of Algebra I and Geometry were derived statistically from the approved cut scores for the other EOC mathematics assessments, Coordinate Algebra and Analytic Geometry.

The final, approved cut scores and associated *impact data* are shown in Table 1. Impact data are the percentage of students who are classified in each achievement level, based on the cut scores. For most tests, the impact data shown in the table reflect the performance of Georgia students in Spring 2015. For Algebra I and Geometry, the impact data are based on Winter 2015.

Table 1. Final, approved cut scores on the final reporting scale and associated impact data

Content	Grade	Approved Cut Scores			Associated Impact Data			
		Dev.	*Prof.*	*Dist.*	*Beg.*	*Dev.*	*Prof.*	*Dist.*
ELA	3	475	525	581	33.1%	30.0%	26.5%	10.4%
	4	475	525	574	29.2%	33.7%	28.0%	9.0%
	5	475	525	587	27.1%	33.7%	31.3%	8.0%
	6	475	525	599	30.5%	30.6%	31.3%	7.7%
	7	475	525	592	30.5%	33.0%	29.9%	6.5%
	8	475	525	581	24.4%	37.4%	31.2%	7.0%
	9TH	475	525	587	23.5%	37.9%	32.4%	6.2%
	AME	475	525	590	25.3%	39.0%	29.5%	6.2%
Math-ematics	3	475	525	580	21.4%	40.6%	30.3%	7.7%
	4	475	525	585	22.6%	38.7%	30.1%	8.5%
	5	475	525	580	25.4%	36.5%	27.5%	10.6%
	6	475	525	580	25.5%	38.9%	26.4%	9.2%
	7	475	525	580	25.2%	37.7%	25.0%	12.2%
	8	475	525	579	24.8%	38.4%	24.8%	12.1%
	COO	475	525	594	30.6%	35.6%	26.1%	7.6%
	ANA	475	525	596	31.9%	35.3%	24.6%	8.2%
	ALG*	475	525	594	24%	40%	30%	6%
	GEO*	475	525	596	19%	24%	30%	27%
Science	3	475	525	566	24.4%	41.3%	26.3%	8.0%
	4	475	525	578	29.0%	37.5%	26.7%	6.8%
	5	475	525	595	29.9%	33.9%	28.4%	7.7%
	6	475	525	610	35.1%	27.1%	31.6%	6.3%
	7	475	525	589	35.1%	29.3%	26.5%	9.1%
	8	475	525	593	38.5%	29.8%	24.8%	6.9%
	PHY	475	525	604	36.3%	32.5%	26.2%	5.0%
	BIO	475	525	609	36.6%	25.1%	29.1%	9.2%
Social Studies	3	475	525	560	26.2%	44.2%	19.7%	10.0%
	4	475	525	570	28.9%	36.4%	25.1%	9.6%
	5	475	525	555	25.3%	45.5%	19.3%	9.8%
	6	475	525	560	29.1%	38.9%	19.5%	12.4%
	7	475	525	564	27.8%	36.2%	21.9%	14.0%
	8	475	525	572	29.5%	37.9%	22.0%	10.6%
	USH	475	525	590	27.2%	33.7%	29.9%	9.2%
	ECO	475	525	610	30.3%	31.1%	31.7%	7.0%

Content	Grade	Approved Cut Scores	Associated Impact Data	
		On-Grade Reading	*Not On-Grade Reading*	*On-Grade Reading*
Reading	3	470	31.9%	68.1%
	4	494	41.1%	58.9%
	5	483	33.7%	66.3%
	6	492	40.2%	59.8%
	[illegible]	472	28.9%	71.1%
	8	482	29.8%	70.2%
	9TH	478	26.3%	73.7%
	AME	482	30.3%	69.7%

** Cut scores for Algebra I and Geometry were derived statistically from the cut scores for Coordinate Algebra and Analytic Geometry. The impact data for these tests are reported from the Winter 2015 administration. This process is described in Section B of this report.*

Executive Summary of the 2018 Georgia Milestones Standard Setting Technical Report

Executive Summary

Summary

- In late 2017 and early 2018, committees of Georgia educators reviewed the cut scores for the Georgia Milestones assessments in science and social studies.
- The original cut scores for these tests were established in 2015 after an extensive *standard setting* process with Georgia educators and stakeholders.
- A *standards validation* was needed in 2017–2018 because of recent changes to the science and social studies content standards (i.e., transition to the *Georgia Standards of Excellence*) and to the tests to align with the GSE.
- Educators made recommendations consistent with the existing cut scores for science and social studies. Accordingly, the validity evidence supports continued use of the cut scores.

Background

On November 7, 2017, and on March 27, 2018, the Georgia Department of Education (GaDOE) partnered with Data Recognition Corporation (DRC) to conduct a standards validation for the Georgia Milestones end-of-course (EOC) assessments of Biology, Physical Science, United States History, and Economics/Business/Free Enterprise; and the end-of-grade (EOG) assessments of science and social studies for grades 5 and 8.

In school year 2017–2018, the Georgia Standards of Excellence (GSE) replaced the Georgia Performance Standards (GPS) as the state's content standards for science and social studies. The Georgia Milestones tests transitioned to the GSE for science and social studies in school year 2017–2018. The test blueprints for these assessments changed as a result of alignment to the GSE.

The test constructs for the eight tests have not changed. Although the GSEs reflect updated language when discussing the tested content (and, for science, an updated organization of that tested content), the underlying knowledge, skills, and abilities measured by the updated tests are consistent with the existing tests.

The purpose of the standards validations was to determine whether the current cut scores for the tests were still valid for continued use on these eight tests. To do so, Georgia educators convened to update the achievement level descriptors (ALDs), to study the cut scores (as applied to the updated tests), and to make recommendations to the GaDOE regarding the suitability of maintaining the current cut scores on the updated tests. This document shows how these educators validated the cut scores through a multi-phased standards validation process.

Standards Validation Methodology

The standards validation was an extension of the original Georgia Milestones standard setting, held in August 2015. The standards validation comprised three phases:

Copyright © 2018 by DRC

1) Georgia educators updated the achievement level descriptors (ALDs) to reflect the language and organization of the Georgia Standards of Excellence (GSE).
2) Two panels of Georgia educators convened in content-based standards validation workshops, where they engaged in a modification of the Bookmark Standard Setting Procedure to review the existing cut scores. These educators reviewed the updated tests, considered the existing cut scores as applied to the new tests, and engaged in content-based conversations about the extent to which the cut scores remained valid for continued use.
3) The GaDOE considered the percentage of students classified in each achievement level on the tests (the *impact data*) to determine whether similar percentages of students would be classified on the tests, in Winter 2017 for EOC and Spring 2018 for EOG, as had been observed on previous administrations of the tests.

During this process, standards validation participants for EOC and EOG made recommendations consistent with the existing cut scores. The GaDOE approved the use of the existing cut scores for the four EOC tests on December 15, 2017, and these cut scores were applied to students' scores during the Winter 2017 administration. In addition, the GaDOE approved the use of the existing EOG cut scores on May 1, 2018, and these cut scores were applied to students' scores during the Spring 2018 administration. GaDOE informed the State Board of Education of the continued use of the existing cut scores for all eight tests.

Achievement Level Descriptors (ALDs)

The ALDs summarize the knowledge, skills, and abilities expected of students in each achievement level. ALDs link students' scores to the GSEs: based on a student's score and achievement level, one can read the ALDs to see a description of the content the student has likely mastered. The ALDs build from *policy ALDs*, shown below, a brief set of descriptors that describe the intention behind each achievement level.

- *Beginning Learners* do not yet demonstrate proficiency in the knowledge and skills necessary at this grade level/course of learning, as specified in Georgia's content standards. The students need substantial academic support to be prepared for the next grade level or course and to be on track for college and career readiness.
- *Developing Learners* demonstrate partial proficiency in the knowledge and skills necessary at this grade level/course of learning, as specified in Georgia's content standards. The students need additional academic support to ensure success in the next grade level or course and to be on track for college and career readiness.
- *Proficient Learners* demonstrate proficiency in the knowledge and skills necessary at this grade level/course of learning, as specified in Georgia's content standards. The students are prepared for the next grade level or course and are on track for college and career readiness.
- *Distinguished Learners* demonstrate advanced proficiency in the knowledge and skills necessary at this grade level/course of learning, as specified in Georgia's content standards. The students are well prepared for the next grade level or course and are well prepared for college and career readiness.

Copyright © 2018 by DRC

Georgia educators drafted the updated ALDs using the GSEs as a base, and the GaDOE worked with DRC to refine them for consistency across tests. An excerpt from the ALDs for EOC U.S. History is presented in Figure 1.

Figure 1. Excerpt from the achievement level descriptors (ALDs) for End-of-Course U.S. History, showing the content-based expectations for students in each achievement level

POLICY ALDS					
ALD	**Standard**	**Beginning Learner**	**Developing Learner**	**Proficient Learner**	**Distinguished Learner**
Policy		Beginning Learners **do not yet demonstrate proficiency in the knowledge and skills** necessary at this grade level/course of learning, as specified in Georgia's content standards. The students need substantial academic support to be prepared for the next grade level or course and to be on track for *college and career readiness* .	Developing Learners **demonstrate partial proficiency in the knowledge and skills** necessary at this grade level/course of learning, as specified in Georgia's content standards. The students need additional academic support to ensure success in the next grade level or course and to be on track for *college and career readiness* .	Proficient Learners **demonstrate proficiency in the knowledge and skills** necessary at this grade level/course of learning, as specified in Georgia's content standards. The students are prepared for the next grade level or course and are on track for *college and career readiness* .	Distinguished Learners **demonstrate advanced proficiency in the knowledge and skills** necessary at this grade level/course of learning, as specified in Georgia's content standards. The students are well prepared for the next grade level or course and are well prepared for *college and career readiness.*
Range ALDS					
		Beginning Learner	**Developing Learner**	**Proficient Learner**	**Distinguished Learner**
Range		A student who achieves at the **Beginning Learner** level demonstrates minimal command of the grade-level standards. The pattern exhibited by student responses indicates that students are most likely able to	A student who achieves at the **Developing Learner** level demonstrates partial command of the grade-level standards. The pattern exhibited by student responses indicates that students are most likely able to	A student who achieves at the **Proficient Learner** level demonstrates proficiency of the grade-level standards. The pattern exhibited by student responses indicates that students are most likely able to	A student who achieves at the **Distinguished Learner** level demonstrates advanced proficiency of the grade-level standards. The pattern exhibited by student responses indicates that students are most likely able to
Colonization through the Constitution					
SSUSH.1 Compare and contrast the development of English settlement and colonization during the 17th Century.					
		recognize that the English settled in North America during the seventeenth century;	describe areas of English settlement in North America during the seventeenth century;	compare and contrast areas of English settlement in North America during the seventeenth century;	compare and contrast areas of English settlement in North America during the seventeenth century, including reasons for settlement, relations with American Indians, and the role of location and place in specific colonial regions;
	USH.1.a	identify mercantilism;	describe trans-Atlantic trade;	investigate how mercantilism and trans-Atlantic trade led to the development of colonies;	
	USH.1.b				
	USH.1.c				
	USH.1.d				

Copyright © 2018 by DRC

Content-Based Standards Validation

Participants engaged in a modification of the Bookmark Standard Setting Procedure to validate the existing cut scores. To do so, participants studied the updated ALDs to learn about the content-based expectations for students in each achievement level. Then, participants saw collections of test items ordered in terms of difficulty. The existing cut scores were presented as *benchmarks* for participants' consideration: participants were asked to consider the knowledge and skills that students would need to demonstrate on the updated tests, as based on the benchmarked (existing) cut scores, and then compare these expectations against the updated ALDs. Participants were instructed to recommend retaining the existing cut scores if there was good correspondence between the benchmarks and the updated ALDs, or to recommend alternative cut scores which would reflect better correspondence.

Participants engaged in two rounds of discussion and iterative judgments. Before the workshops, it was hypothesized that participants would recommend cut scores which were similar to, but not exactly equal to, the existing cut scores. The rationale behind this hypothesis was that nearly any group of educators going through an iterative, judgmental process like the Bookmark Procedure tend to arrive at slightly different judgments at the end of the process. Slight differences in cut score recommendations could be attributed to random statistical errors. This hypothesis was later used to inform the interpretation of the workshop results, presented under the heading "Review of Recommendations."

Table 1 shows the median recommended cut scores and associated impact data from the content-based standards validations. Impact data are the percentages of students who would be classified in each achievement level on the Spring 2017 assessments if these median cut scores were applied.

Table 1. Recommended Cut Scores and Associated Impact Data for EOC and EOG Spring 2017

	Median Cut Score Recommendations			% Students by Level Based on Recommended Cuts			
Grade/ Course	***Developing Learner***	***Proficient Learner***	***Distinguished Learner***	***Beginning Learner***	***Developing Learner***	***Proficient Learner***	***Distinguished Learner***
BIOL	478	523	603	32.3%	20.9%	32.4%	14.3%
PHSC	472	519	604	29.1%	27.9%	34.2%	8.8%
HIST	475	525	585	24.3%	29.9%	30.5%	15.2%
ECON	470	520	610	23.8%	26.8%	36.7%	12.6%
SC G5	474	526	595	30.5%	31.4%	28.3%	9.8%
SC G8	474	525	590	39.4%	31.4%	23.7%	5.5%
SS G5	475	525	555	26.5%	44.0%	19.0%	10.4%
SS G8	475	521	569	25.1%	32.2%	28.6%	14.2%

Review of the Recommendations

As hypothesized, educators at the content-based standards validation workshops recommended cut scores that were similar to the existing cut scores. GaDOE worked with DRC to evaluate the recommendations in context. Table 2 shows the difference between the median cut score recommendations and the existing cut scores, expressed in multiples of the conditional standard error of measurement (CSEM). The CSEM quantifies the amount of statistical error associated with the test. If a student were tested many times, one would expect her scores to fall within a range of ±1.0 CSEM about 2/3 of the time.

Copyright © 2018 by DRC

Table 2. Comparison of Median Cut Score Recommendations and Existing Cut Scores, with Differences Expressed in Terms of Conditional Standard Error of Measurement (CSEM)

	Median Cut Scores Recomm.			Existing Cut Scores			Difference in CSEM Units		
Grade/ Course	***Dev Learner***	***Prof Learner***	***Dist Learner***	***Dev Learner***	***Prof Learner***	***Dist Learner***	***Dev Learner***	***Prof Learner***	***Dist Learner***
BIOL	478	523	603	475	525	609	+0.17	-0.11	-0.24
PHSC	472	519	604	475	525	604	-0.18	-0.33	-0-
HIST	475	525	585	475	525	590	-0-	-0-	-0.26
ECON	470	520	610	475	525	610	-0.28	-0.26	-0-
SC G5	474	526	595	475	525	595	-0.06	+0.06	-0-
SC G8	474	525	590	475	525	593	-0.06	-0-	-0.14
SS G5	475	525	555	475	525	555	-0-	-0-	-0-
SS G8	475	521	569	475	525	572	-0-	-0.31	-0.18

The GaDOE considered the recommendations made by members of the standards validation committees and the existing cut scores. Working with DRC, GaDOE made four primary findings:

1) The content-based expectations for students in each achievement level have not changed significantly since 2015. The underlying expectations for student achievement in each achievement level have not changed, although the content standards and ALDs have been updated, especially in terms of wording and organization.

2) Participants' bookmark and cut score recommendations were very similar to those from the 2015 standard setting committee. The final-round recommendations of participants at the standards validations were very similar to those from the original 2015 standard setting.

3) The median cut score recommendations were all very close to the existing cut scores, to the point of being statistically indistinguishable. As shown in Table 2, the average difference from the existing cut scores was -0.09 CSEM, and all were within a range of ±0.33 CSEM. Within this narrow range, it is difficult to argue that scale scores are significantly different.

4) The impact data observed in Winter 2017 for EOC and Spring 2018 for EOG were similar to those from previous administrations of the tests when the existing cut scores were applied. This similarity supports the contention that the expectations for students in each achievement level has not changed, and that the existing cut scores are valid for continued use.

GaDOE considered the differences between the median recommendations made at the standards validation workshops and the cut scores established at the 2015 standard setting, and it weighted these very small differences against the relative costs to the testing system of changing cut scores (e.g., impairing existing longitudinal studies).

Based on the totality of the evidence, the GaDOE determined that there were no significant differences between the new science and social studies assessments and the content assessed by the prior assessments; and that the differences between the judgments made at the 2017 and 2018 standards validation workshops and the original 2015 cut scores were not statistically different. That is, the recommendations made by Georgia educators during the standards validation process were consistent

Copyright © 2018 by DRC

with the existing cut scores, and the validity evidence collected during this process supports the continued use of the cut scores.

Table 3 shows the final approved cut scores and associated impact data for EOC using Winter 2017 administration and for EOG using Spring 2018 administration data.

Table 3. Validated Cut Scores and Associated Impact Data for EOC Winter 2017, EOG Spring 2018

	Validated Cut Scores			% Students by Level Based on Validated Cut Scores			
Grade/ Course	***Developing Learner***	***Proficient Learner***	***Distinguished Learner***	***Beginning Learner***	***Developing Learner***	***Proficient Learner***	***Distinguished Learner***
BIOL	475	525	609	31.0%	25.3%	31.8%	11.9%
PHSC	475	525	604	38.0%	35.0%	23.4%	3.7%
HIST	475	525	590	28.8%	30.5%	30.9%	9.8%
ECON	475	525	610	26.0%	28.3%	33.4%	12.3%
SC G5	475	525	595	29.6%	30.0%	28.7%	11.7%
SC G8	475	525	593	38.3%	30.7%	24.2%	6.9%
SS G5	475	525	555	22.9%	46.3%	20.2%	10.6%
SS G8	475	525	572	22.1%	36.5%	28.3%	13.1%

Copyright © 2018 by DRC